

2014

Computational Insights into the Accuracy and Editing of Aminoacyl tRNA Synthetases

Grant B. Fortowsky
University of Windsor

Follow this and additional works at: <http://scholar.uwindsor.ca/etd>

Recommended Citation

Fortowsky, Grant B., "Computational Insights into the Accuracy and Editing of Aminoacyl tRNA Synthetases" (2014). *Electronic Theses and Dissertations*. Paper 5134.

This online database contains the full-text of PhD dissertations and Masters' theses of University of Windsor students from 1954 forward. These documents are made available for personal study and research purposes only, in accordance with the Canadian Copyright Act and the Creative Commons license—CC BY-NC-ND (Attribution, Non-Commercial, No Derivative Works). Under this license, works must always be attributed to the copyright holder (original author), cannot be used for any commercial purposes, and may not be altered. Any other use would require the permission of the copyright holder. Students may inquire about withdrawing their dissertation and/or thesis from this database. For additional inquiries, please contact the repository administrator via email (scholarship@uwindsor.ca) or by telephone at 519-253-3000ext. 3208.

Computational Insights into the Accuracy and Editing of Aminoacyl tRNA Synthetases

By

Grant B. Fortowsky

A Thesis

Submitted to the Faculty of Graduate Studies
through the Department of Chemistry and Biochemistry
in Partial Fulfillment of the Requirements for
the Degree of Master of Science
at the University of Windsor

Windsor, Ontario

2014

© 2014 Grant B. Fortowsky

Computational Insights into the Accuracy and Editing of Aminoacyl tRNA Synthetases

by

Grant Fortowsky

APPROVED BY:

L. Porter
Department of Biology

H. Eichhorn
Department of Chemistry & Biochemistry

J. Gauld, Advisor
Department of Chemistry & Biochemistry

June 12, 2014

Declaration of Co-Authorship

I hereby declare that this thesis incorporates material that is a result of joint research as follows:

Chapter 3: was done in collaboration with Daniel J. Simard and under the supervision of Prof. James W. Gauld.

Chapter 4: was done in collaboration with Daniel J. Simard and under the supervision of Prof. James W. Gauld.

Chapter 5: was done in collaboration with Daniel J. Simard and under the supervision of Prof. James W. Gauld.

I certify that, to the best of my knowledge, my thesis does not infringe upon anyone's copyright nor violate any proprietary rights and that any ideas, techniques, quotations, or any other material from the work of other people included in my thesis, published or otherwise, are fully acknowledged in accordance with the standard referencing practices. Furthermore, to the extent that I have included copyrighted material that surpasses the bounds of fair dealing within the meaning of the Canada Copyright Act, I certify that I have obtained a written permission from the copyright owner(s) to include such material(s) in my thesis and have included copies of such copyright clearances to my appendix.

I declare that this is a true copy of my thesis, including any final revisions, as approved by my thesis committee and the Graduate Studies office, and that this thesis has not been submitted for a higher degree to any other University or Institution.

Abstract

Proteins are ubiquitous in nature, carrying out many of life's functions. However, for these biomolecules to function correctly they must be accurately synthesized. The task of ensuring the desired accuracy is achieved falls to Aminoacyl tRNA Synthetases (aaRS). These enzymes catalyze the activation and transfer of amino acids to tRNA. In addition to carrying out these reactions, they also edit against the incorporation of incorrect amino acids. Within this thesis the many facets of aaRS editing and accuracy are examined in detail. In particular, the editing mechanism of Hcy and Hse in MetRS is elucidated in detail. These results allow us to infer generalizations of the editing of Hcy within many aaRS's. Zn(II)'s role in catalysis and accuracy for SerRS, CysRS and ThrRS is also examined in some detail and gives key insights into a new potential role it may play.

Dedication

I dedicate this work to my family. Thank you so much for all your encouragement. I would also like to single out my little brother, you have always been nice enough to cheer me up when I am down and rude enough to knock me back down when I'm feeling up. Thanks for being there for me!

Acknowledgments

Throughout the last two years I have had the pleasure of working with many great people and made many friends along the way. I would like to thank Dr. James W. Gauld, my supervisor, for allowing me the great opportunity to pursue graduate studies. With his guidance I have been able to overcome all of the difficulties that arose during the course of my graduate studies. This has allowed me to become a more diplomatic and effective researcher.

I would also like to thank my committee, Prof. Eichhorn and Prof. Porter for taking the time to read this thesis. In addition I would like to thank Prof. Eichhorn for helping me rationalize different organic reactions that pertained to my studies, your insights proved to be very valuable.

The Gauld group past and present has also provided me with countless hours of entertainment as well as a great deal of knowledge and help conducting research. For these reasons I am thrilled to thank Dr. Wenjuan Huang, Dr. Eric Buchnell, Bogdan Ion, Hisham Dokainish, Natalia Mroz, Rami Gherib, Erum Kazim, Phil De Luna, Daniel Simard, Mohamed Aboelnga, Anthony Deschamps and Peter Boateng. Without these fine people my masters studies would have been a great deal more difficult and a lot less enjoyable.

My most heartfelt thanks goes out to my little blue eyed brother. Without your company and entertainment at home I am not sure that my sanity would have

survived these past two years. In particular I always smile thinking about how cool some stairs are or how Duncan always wants to dive again. Thanks for the entertainment!

Table of Contents

| | |
|--|-------|
| Declaration of Co-authorship | III |
| Abstract..... | IIV |
| Dedication..... | V |
| Acknowledgments..... | VI |
| List of Figures..... | XII |
| List of Tables..... | XV |
| List of Schemes | XVII |
| List of Appendixes | XVIII |
| List of Abbreviations | XX |
| Chapter 1: Introduction | 1 |
| 1.1 Introduction | 2 |
| 1.2 Proteins | 2 |
| 1.3 Enzymes | 3 |
| 1.4 Central Dogma..... | 4 |
| 1.5 tRNA..... | 5 |
| 1.6 Aminoacyl tRNA synthetases..... | 6 |
| 1.6.1 Role in protein synthesis..... | 6 |
| 1.6.2 Accuracy and editing..... | 7 |
| 1.6.3 Health effects..... | 8 |
| 1.7 Computational Chemistry..... | 9 |
| 1.8 References..... | 10 |

| | |
|--|-----------|
| Chapter 2: Theoretical Methods | 13 |
| 2.1 Introduction | 14 |
| 2.2 Classical Mechanics | 14 |
| 2.2.1 Molecular Mechanics | 15 |
| 2.2.2 Molecular Dynamics..... | 15 |
| 2.2.3 Uses and Limitations..... | 16 |
| 2.3 Quantum Mechanics..... | 17 |
| 2.4 The Schrödinger Equation | 17 |
| 2.5 Born-Oppenheimer Approximation..... | 18 |
| 2.6 The Orbital Approximation..... | 19 |
| 2.7 Basis Set Expansion..... | 20 |
| 2.8 The Variational Theorem..... | 21 |
| 2.9 Hartree Fock Theory and Wave Function Based Methods..... | 22 |
| 2.10 Density Functional Theory | 23 |
| 2.10.1 Hohenberg-Kohn Theorem | 23 |
| 2.10.2 Kohn-Sham Theorem | 24 |
| 2.11 Basis Sets..... | 26 |
| 2.11.1 Split Valence Basis Sets | 26 |
| 2.11.2 Polarization Basis sets..... | 27 |
| 2.11.3 Diffuse Functions | 27 |
| 2.12 Solvation..... | 27 |
| 2.13 Quantum Mechanics/Molecular Mechanics..... | 28 |
| 2.14 Potential Energy Surfaces (PES) | 30 |
| 2.15 Technical Aspects and Units | 31 |

| | |
|-----------------------|----|
| 2.16 References | 31 |
|-----------------------|----|

Chapter 3: Enzymatic and Substrate-Assisted Editing of Highly Reactive

| | |
|---|-----------|
| Nonstandard Amino Acids by Aminoacyl-tRNA Synthetases..... | 34 |
| 3.1 Introduction | 35 |
| 3.2 Theoretical Methods..... | 36 |
| 3.3 Results and Discussion | 37 |
| 3.4 Conclusions..... | 43 |
| 3.5 References..... | 43 |

Chapter 4: Mechanistic Insights into the Self-Cyclization of Homocysteine and

| | |
|---|-----------|
| Homosereine | 47 |
| 4.1 Introduction | 48 |
| 4.2 Theoretical Methods..... | 48 |
| 4.3 Phosphate facilitated cyclization in MetRS..... | 49 |
| 4.4 An alternative the base in the self-cyclization of Hcy..... | 51 |
| 4.5 Mechanism of Hcy cyclization | 53 |
| 4.6 Hse Self-Cyclization..... | 56 |
| 4.7 A General Motif of Hcy Editing | 59 |
| 4.8 Conclusions..... | 62 |
| 4.9 References..... | 63 |

Chapter 5: A Computational Study into an Additional Key Role of Zinc in SerRS:

| | |
|---|-----------|
| Inhibition of Substrate Self-Cyclization | 65 |
|---|-----------|

| | |
|---|--------|
| 5.1 Introduction | 66 |
| 5.2 Computational Methods..... | 70 |
| 5.2.1 Molecular Dynamics..... | 70 |
| 5.2.2 QM/MM Calculations | 70 |
| 5.2.3 Model Preparation | 71 |
| 5.2.4 Quantum Mechanical Cluster..... | 72 |
| 5.3 Results and Discussion..... | 72 |
| 5.3.1 Quantum Mechanical Cluster Kinetics/Thermodynamics..... | 72 |
| 5.3.2 QM/MM Model Choice | 76 |
| 5.3.3 QM/MM without Zn(II)..... | 76 |
| 5.3.4 QM/MM without Zn(II) and fully protonated α -amine | 78 |
| 5.3.5 QM/MM of Native SerRS | 80 |
| 5.4 Applications of Current Work | 83 |
| 5.5 Conclusions..... | 86 |
| 5.6 References..... | 87 |
| Chapter 6: Conclusions | 91 |
| 6.1 Conclusions..... | 92 |
| 6.2 References..... | 94 |
| Vita Auctoris | 95 |

List of Figures

| | |
|--|----|
| Figure 1.1. An illustrated representation of a typical peptide bond found in biochemical systems. | 2 |
| Figure 1.2. A comparison of the similarity between Serine on the left and Cysteine on the right, highlighting the fact that there is only one atom different between the two. | 4 |
| Figure 1.3. A schematic representation of the central dogma, which details flow of genetic information to proteins in biological systems. | 5 |
| Figure 2.1. An illustration of a QM/MM model in which the environment is modeled at the MM level of theory while the chemically active center uses a QM level of theory. | 29 |
| Figure 2.2. A schematic representation of a potential energy surface (PES), where A is the reactant complex, B and D are transition states, C is an intermediate and E is the product complex. | 30 |
| Figure 3.1. The proposed editing mechanism of Hcy in MetRS by which a thiol subsite facilitates self-cyclization..... | 36 |
| Figure 3.2. Overlaid average MetRS active site-bound structures of Met-AMP, Hcy-AMP and Hse-AMP. | 37 |

| | |
|--|----|
| Figure 3.3. Free Energy surfaces (kJ mol^{-1}) for "editing" of Hcy-AMP in MetRS using as the mechanistic base (A) a Hcy-AMP phosphate non-bridging oxygen, or (B) Asp259. | 38 |
| Figure 3.4. An illustration of the conformation of Hcy-AMP when bound in the active site of IleRS (left) and MetRS (right). | 43 |
| Figure 4.1. Molecular dynamics simulation of MetRS showing the distance between the phosphates non bridging oxygen and Hcy's sulfur with respect to time. | 50 |
| Figure 4.2. A schematic representation of the distances calculated in the phosphate base mediated self-cyclization of Hcy. | 51 |
| Figure 4.3. A graph showing the distance between the Asp259 and the Hcy's carbonyl carbon with respect to time during the molecular dynamics simulation of MetRS. | 52 |
| Figure 4.4. A graph showing the distance between the Asp259 and the Hcy's sulfur with respect to time during the molecular dynamics simulation of MetRS. | 53 |
| Figure 4.5. The potential energy surface for the cyclization reaction of Hcy-AMP in MetRS with all relative energies reported in kJ mol^{-1} . | 54 |
| Figure 4.6. A) A graph showing the distance between the oxygen of Asp259 base and the Hse's oxygen during the molecular dynamics simulation. (At approximately 5 ns the carboxylate undergoes a rotation exchanging oxygen positions.) B) A graph showing the distance between the carbon of Asp259 base and the Hse's oxygen during the molecular dynamics simulation of MetRS. | 57 |
| Figure 4.7. The potential energy surface for the cyclization reaction of Hse-AMP in MetRS with all relative energies reported in kJ mol^{-1} . | 59 |

| | |
|---|----|
| Figure 4.8. An illustration of Hcy and a carboxylate (either aspartic acid or glutamic acid) within the active site. Selected distances are highlighted A) $S_{\text{Hcy}}-O_{\text{carboxylate}}$ B) $S_{\text{Hcy}}-O_{\text{phos}}$ C) $O_{\text{carboxylate}}-C_{\text{sp}2}$ D) $S_{\text{Hcy}}-C_{\text{sp}2}$ | 61 |
| Figure 5.1. The rate-limiting step of the self cyclization reaction for cysteine, threonine and serine using IEF-PCM-B3LYP/6-311G(2df,p)//IEF-PCM-B3LYP/6-31G(d,p)+ ΔE_{Gibbs} | 73 |
| Figure 5.2. The bond distance with respect to dielectric of the attacking hydrogen from the hydroxyl group and the accepting phosphate's oxygen for the rate limiting transition state. | 74 |
| Figure 5.3. The overall thermodynamics of the self cyclization reaction for cysteine, threonine and serine using IEF-PCM-B3LYP/6-311G(2df,p)//IEF-PCM-B3LYP/6-31G(d,p)+ ΔE_{Gibbs} .. | 75 |
| Figure 5.4. A) Optimized Reactive complex with selected bond distances (angstroms) highlighted. B) Optimized first transition state with selected bond distances (angstroms) highlighted. C) Optimized intermediate with selected bond distances (angstroms) highlighted. D) Optimized second transition state with selected bond distances (angstroms) highlighted..... | 77 |
| Figure 5.5. Calculated PES for the self cyclization reaction without Zn(II) and an neutral amine, B3LYP energies shown in kJ mol^{-1} .. | 78 |
| Figure 5.6. A) Optimized Reactive complex with selected bond distances (angstroms) highlighted. B) Optimized first transition state with selected bond distances (angstroms) highlighted. C) Optimized intermediate with selected bond distances (angstroms) highlighted. D) Optimized second transition state with selected bond distances (angstroms) highlighted..... | 79 |

| | |
|---|----|
| Figure 5.7. Calculated PES for the self cyclization reaction without Zn(II) and an protonated amine, B3LYP energies shown in kJ mol^{-1} | 80 |
| Figure 5.8. A) Optimized Reactive complex with selected bond distances (angstroms) highlighted. B) Optimized first transition state with selected bond distances (angstroms) highlighted. C) Optimized intermediate with selected bond distances (angstroms) highlighted. D) Optimized second transition state with selected bond distances (angstroms) highlighted..... | 81 |
| Figure 5.9. Calculated PES for the self cyclization reaction in the native enzyme, B3LYP energies shown in kJ mol^{-1} | 83 |
| Figure 5.10. The rate-limiting step of the self cyclization reaction for glutamate phosphate using IEF-PCM-B3LYP/6-311G(2df,p)//IEF-PCM-B3LYP/6-31G(d,p)+ ΔE_{Gibbs} , IEF-PCM-M06/6-311G(2df,p)//IEF-PCM-B3LYP/6-31G(d,p)+ ΔE_{Gibbs} , IEF-PCM-M06-2X/6-311G(2df,p)//IEF-PCM-B3LYP/6-31G(d,p)+ ΔE_{Gibbs} | 83 |
| Figure 5.11. The overall thermodynamics of the self cyclization reaction for glutamate phosphate using IEF-PCM-B3LYP/6-311G(2df,p)//IEF-PCM-B3LYP/6-31G(d,p)+ ΔE_{Gibbs} , IEF-PCM-M06/6-311G(2df,p)// IEF-PCM-B3LYP/6-31G(d,p)+ ΔE_{Gibbs} , IEF-PCM-M06-2X/6-311G(2df,p)//IEF-PCM-B3LYP/6-31G(d,p)+ ΔE_{Gibbs} ,... | 83 |

List of Tables

Table 3.1. Average distance (Angstroms) obtained from MD simulations (see text) between C_{carb} of the substrate and nearest carboxyl oxygen of acidic active site residues..... 42

Table 4.1. An overview of selected average distances for Hcy bound in MetRS, LeuRS, ValRS and IleRS..... 62

List of Schemes

| | |
|---|----|
| Scheme 1.1. A schematic illustration of ATP's attack on an amino acid in order to activate it in the first reaction carried out by aaRSs. | 7 |
| Scheme 1.2. An overview of the transfer step of aaRS's in which the A76 of tRNA attacks the activated amino acid cleaving AMP..... | 7 |
| Scheme 1.3. A schematic representation of Hcy editing in MetRS followed by a detrimental attack on lysine by the thiolactone product..... | 9 |
| Scheme 4.1. An illustration a reaction pathway for the cyclization of Hcy in MetRS utilizing Asp259 as a mechanistic base..... | 54 |
| Scheme 4.2. An illustration a reaction pathway for the cyclization of Hse in MetRS utilizing Asp259 as a mechanistic base..... | 58 |
| Scheme 5.1. A schematic illustration of a) the activation of amino acids and b) the transfer of amino acids to their cognate tRNA in protein synthesis. | 66 |
| Scheme 5.2. A schematic illustration highlighting the self-cyclization editing function in MetRS..... | 68 |
| Scheme 5.3. An illustration of possible side reactions that could take place in SerRS, ThrRS and CysRS. | 69 |
| Scheme 5.4. The mechanism by which glutamine synthetase converts glutamate to glutamine..... | 84 |
| Scheme 5.5. A possible self-cyclization reaction that may occur in the active site... .. | 84 |

Appendixes

For access to appendixes see attached CD.

Table A1. Optimized xyz coordinates for the high layer of all QM/MM models examined in chapter 3 and 4.

Figure B1. The rate-limiting step of the self cyclization reaction for Cysteine, Threonine and Serine using IEF-PCM-M06/6-311G(2df,p)// IEF-PCM-B3LYP/6-31G(d,p)+ ΔE_{Gibbs}

Figure B2. The rate-limiting step of the self cyclization reaction for Cysteine, Threonine and Serine using IEF-PCM-M06-2X/6-311G(2df,p)// IEF-PCM-B3LYP/6-31G(d,p)+ ΔE_{Gibbs}

Figure B3. The overall thermodynamics of the self cyclization reaction for Cysteine, Threonine and Serine using IEF-PCM-M06/6-311G(2df,p)// IEF-PCM-B3LYP/6-31G(d,p)+ ΔE_{Gibbs}

Figure B4. The overall thermodynamics of the self cyclization reaction for Cysteine, Threonine and Serine using IEF-PCM-M06-2X/6-311G(2df,p)// IEF-PCM-B3LYP/6-31G(d,p)+ ΔE_{Gibbs}

Figure B5. Calculated PES for the self cyclization reaction without Zn(II) and an neutral amine, B3LYP energies shown in blue, M06 shown in red and M06-2X shown in green.

Figure B6. Calculated PES for the self cyclization reaction without Zn(II) and an protonated amine, B3LYP energies shown in blue, M06 shown in red and M06-2X shown in green.

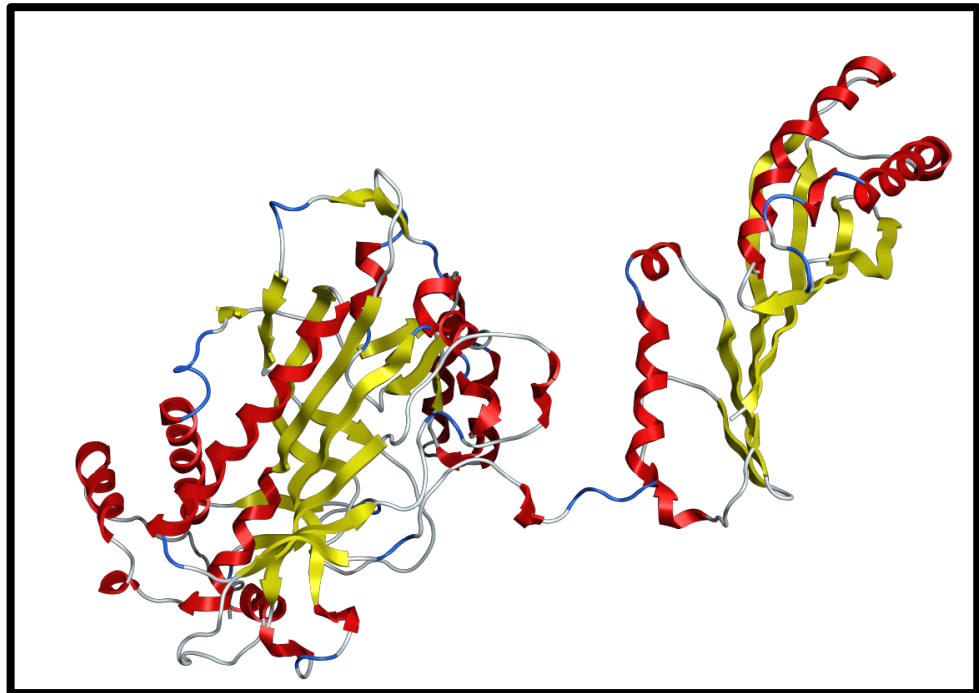
Figure B7. Calculated PES for the self cyclization reaction in the native enzyme, B3LYP energies shown in blue, M06 shown in red and M06-2X shown in green.

List of Abbreviated Symbols

| | |
|-------|---------------------------------------|
| aa | amino acid |
| aaAMP | aminoacyl adenylate |
| aaRS | aminoacyl-tRNA synthetase |
| DFT | density dunctional theory |
| GGA | generalized-gradient approximation |
| GTF | Gaussian type funtions |
| Hcy | homocysteine |
| Hse | homosereine |
| LCAO | linear combination of atomic orbitals |
| LDA | local density approximation |
| MD | molecular dynamics |
| ME | mechanical embedding |
| MM | molecular mechanics |
| MO | molecular orbitals |
| MOE | Molecular Operating Environment |
| PCM | polarized continuum method |
| PDB | protein data bank |
| PES | potential energy surfaces |

| | |
|-------|---------------------------------------|
| QM | quantum mechanics |
| QM/MM | quantum mechanics/molecular mechanics |
| RMSD | root mean square deviation |
| RNA | ribonucleic acids |
| STO | Slater type orbitals |
| tRNA | transfer-RNA |
| UroD | Uroporphyrinogen decarboxylase |

Chapter 1



Introduction

1.1 Introduction

It has been suggested that early in the evolution of life ribonucleic acids (RNA) were responsible for both encoding the genetic information of cells and were cellular biocatalysts, catalyzing required cellular reactions.¹ This catalytic ability of RNA (in addition to genetic encoding) still exists and is exemplified by ribozymes; catalytic RNA.² For example, the hepatitis delta virus (a human pathogen) uses ribozymes for viral replication and amongst ribozymes and enzymes is the fastest known self cleaving biomolecule!³ However over billions of years of evolution, RNA has taken a back seat to other biomolecules for a number of life's functions. For instance, genetic information is now encoded within DNA in higher order organisms while proteins are now the predominate type of cellular biocatalyst. This shift has resulted in proteins now playing a significant role in many of life's biochemical processes.

1.2 Proteins

Proteins are biological polymers consisting of amino acids linked together by peptide bonds (amide bonds), as shown in figure 1.1.

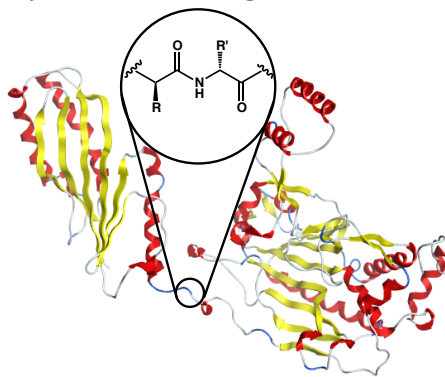


Figure 1.1. An illustrated representation of a typical peptide bond found in biochemical systems.

These biomolecules, made from deceptively simple subunits, are involved in almost every aspect of life. For example, they are involved in structural roles; forming hair and nails of animals along with key components of connective tissue such as collagen and elastin.² In addition, proteins are also instrumental in the transportation of molecules. A famous example of this is hemoglobin; which has the important job of transporting oxygen from our lungs to every other part of our body.² Signaling is another important role that proteins undertake. G-proteins are one such protein and can be activated by a variety of molecules such as epinephrine.² Proteins can also help to speed up reactions that are necessary for life; these proteins are collectively known as enzymes.^{2,4}

1.3 Enzymes

These catalytic proteins are amongst the most important biomolecules for life. Uroporphyrinogen decarboxylase (UroD) exemplifies this importance, as it is one of the most efficient enzymes known.⁵ In heme synthesis the decarboxylation of the 4 acetates of UroD is painstakingly slow under standard conditions, having a half-life of 2.3 billion years for the reaction. This rate is much too slow to support life as we know it, however UroD manages to speed up this process by a factor of 1.2×10^{17} .⁶ This outstanding rate enhancement (albeit not necessarily this large) is typical of many enzymes, not just UroD, and is only one of many reasons why enzymes are studied.

Many enzymes have an intriguing ability to distinguish between similar molecules. Aminoacyl tRNA synthetases (aaRSs) exemplify this ability. This class of enzymes are able to distinguish between cysteine and serine, which differ in only one atom (figure 1.2), by a factor of 10^8 .⁷

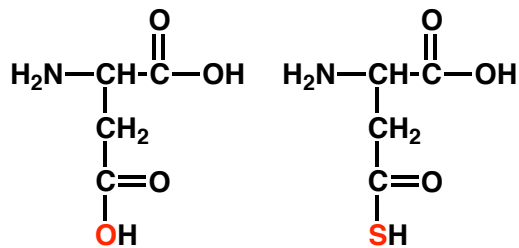


Figure 1.2. A comparison of the similarity between serine on the left and cysteine on the right, highlighting the fact that there is only one atom different between the two.

The remarkable abilities of enzymes have led biological systems to rely on them for the assembly of virtually all biomolecules. This reliance on enzymes for synthesis of biological structures including other enzymes brings up an intriguing question of their origin. This paradox is not easily answered by chemistry alone therefore for an answer we must examine the central dogma of molecular biology.

1.4 Central Dogma

It is now common knowledge that DNA contains our genetic information. It is, however, less well-known how our body makes use of this information. First the information must be transported to the ribosomes, which assemble proteins. Unfortunately, DNA cannot make this journey. Instead it must be transcribed into mRNA which is then transported to the ribosomes. This movement of genetic information to proteins is referred to as the "Central Dogma" (figure 1.3).²

More gennerally, it describes the net transfer of information from genetic material to proteins. However, it leaves out how the right amino acid knows when to assemble into a protein. This is a complex matter since there are 20 natural amino acids; meaning a simple protein consisting of five amino acids has 3,200,000 different ways it could be assembled. The huge amount of variation might cause you to believe that

errors in protein synthesis are common. However, in general this is not in fact the case.^{8,9}

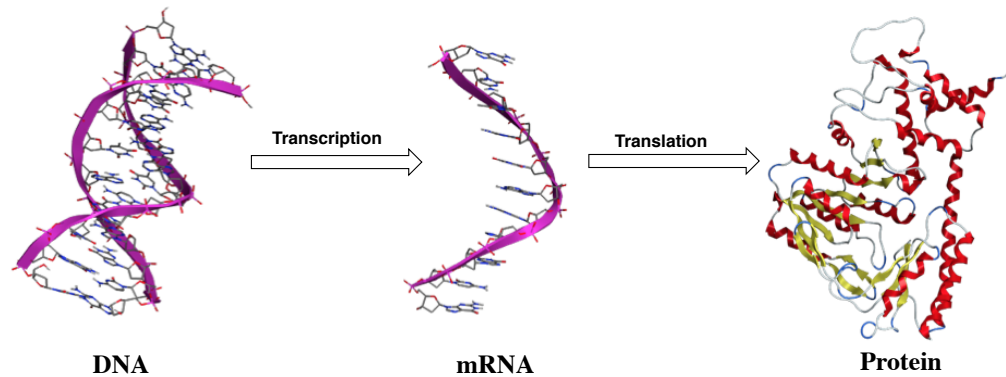


Figure 1.3. A schematic representation of the central dogma, which details flow of genetic information to proteins in biological systems.

1.5 Transfer RNA (tRNA)

Amino acids are transported to the ribosomes by a special version of RNA called transfer-RNA (tRNA). Overall this molecule has a characteristic "L" shape tertiary structure, however it is often more useful to view it from its secondary structure which resembles a cloverleaf.^{10,11} From this view it is easy to see the anticodon which will bind to the codons of mRNA and the acceptor stem which is key for recognition.^{10,11} The actual transfer of amino acids is accomplished by attaching amino acids to the terminal ribose group of tRNA.⁸ A consequence of this is that for every genetically encoded amino acid there is at least one corresponding tRNA molecule that transports it to the ribosome, where it is then incorporated into proteins.⁸ An interesting consequence of every amino acid having its own tRNA, is that there must exist a specific enzyme that recognizes the right amino acid and tRNA and is then able able to link them.

1.6 Aminoacyl tRNA Synthetases

Aminoacyl tRNA synthetases (aaRSs) are a class of enzymes that are crucial for protein synthesis.⁸ They can be divided into two sub-classes: "class 1" and "class 2". They differ in many aspects, such as which hydroxyl will attack the carbonyl carbon, 2' in class 1 and 3' in class 2.^{8,12} Class 1 also bind tRNA along its minor groove with a Rosman fold motif and will consequently bind ATP in a linear fashion.^{8,12} In contrast class 2 binds tRNA in the major groove using anti parallel β -sheets flanked by α -helices and bind ATP in a bent conformation.^{8,12}

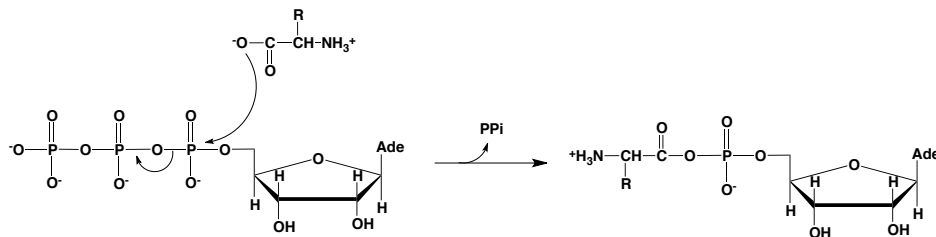
In addition to their key roles in protein synthesis, aaRSs are important for a number of other cellular processes.¹³ It is also noted that, in part because of their ancient origins, they can be useful for tracing evolution and evolutionary relationships.^{14,15}

1.6.1 Role in Protein Synthesis

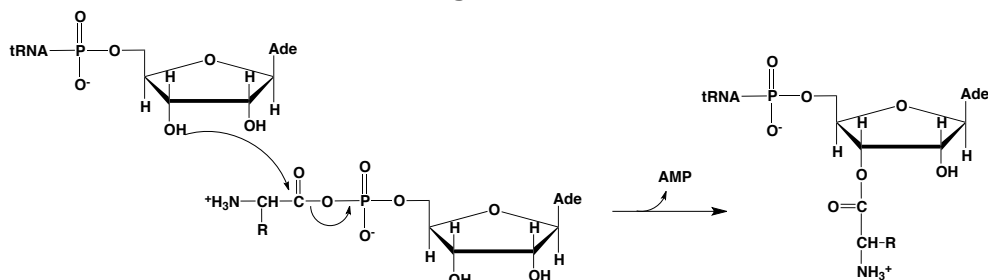
As mentioned, aaRSs are instrumental in the process of protein synthesis. What makes these enzymes even more impressive is that they catalyze multiple reactions within the same active site.¹⁶ First, they must bind their cognate amino acid within their active site as well as ATP. These are then reacted to form an aminoacyl-adenylate (aa-AMP) as shown in scheme 1.1.

After activation, the aminoacyl moiety of the aa-AMP must be transferred to the corresponding cognate tRNA molecule. For this to be done, either the 3' or 2' hydroxyl (depending on the class of aaRS) will attack the aminoacyl moiety's carbonyl carbon; giving the aminoacylated tRNA (aa-tRNA) and AMP (scheme 1.2).^{8,12}

Scheme 1.1. A schematic illustration of ATP's attack on an amino acid in order to activate it in the first reaction carried out by aaRSs.



Scheme 1.2. An overview of the transfer step of aaRS's in which the A76 of tRNA attacks the activated amino acid cleaving AMP.



The aa-tRNA is then released from the aaRS. In general, aaRSs are able to attach the right amino acid to the corresponding cognate tRNA. Unfortunately, however, this is not always the case.¹⁷⁻¹⁹

1.6.2 Accuracy and Editing

AaRSs are generally held to be outstandingly accurate. Indeed, it has been stated that on average they only make 1 error in every 10,000 reactions.⁸ For example CysRS is able to discriminate Cys from Ser, which differ by only a single atom (Cys contains a sulfur where Ser contains an oxygen), by a factor of 10^8 .²⁰ Similarly, SerRS discriminates Ser from Thr; the latter containing a methyl group in its side chain where Ser contains a hydrogen.²¹

One of the ways in which this specificity is accomplished is by utilizing binding site, that only allow for the correct amino acid to bind.^{17,21,22} An example of this is CysRS whose active site contains a Zn(II) center, thus allowing cysteine to bind through a strong thiol-zinc interaction.²⁰ This interaction is only present in the case of cysteine and therefore allows for discrimination by binding alone. It is noted that SerRS also achieves its discrimination through binding.²¹

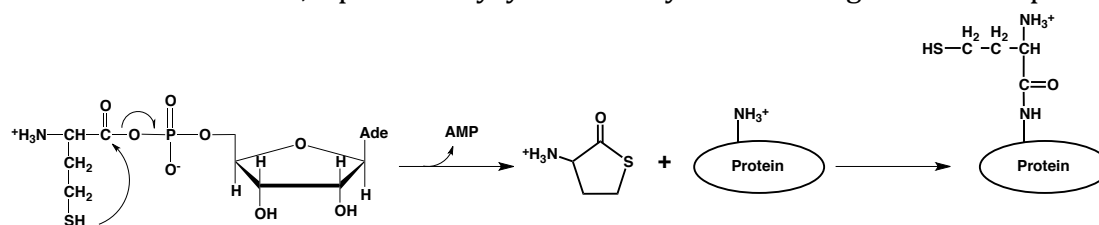
Unfortunately, the solution to achieving the required accuracy for protein synthesis cannot always easily be achieved in all aaRSs. In ThrRS both Thr and Ser are readily activated and transferred at similar rates onto the cognate tRNA for Thr. As a result, other methods must be employed.²³⁻²⁶ One such approach is to use a separate editing site whose role is to hydrolyze the aminoacyl-tRNA bond in the case of non-cognate amino acids bound to the tRNA moiety.²⁷ This process, known as post-transfer editing, ensures that an incorrect amino acid is removed before it can be incorporated into a protein.⁹ Alternatively, an aaRS may exploit pre-transfer editing. In this process the aminoacyl-AMP bond is hydrolyzed before the aminoacyl moiety can be transferred to the tRNA.⁹ It is by this very method that aaRSs ensure that the highly reactive Hcy is never incorporated into proteins. It is noted that if this fidelity and accuracy of aminoacylation is not achieved a number of health effects could result.

1.6.3 Health Effects

The mis-functioning of aaRSs have been associated with many illnesses ranging from HIV to different types of cancers and atherosclerosis.²⁸⁻³⁰ Perhaps even more interesting is how they are thought to be linked to these illnesses. For example, in the case of HIV it is believed that the interaction of LysRS with the HIV capsid enables gene expression to occur.³⁰ In the case of cancer, while the specific roles of aaRS's is

less clear, they are thought to be involved in signaling.²⁹ Meanwhile, cardiovascular diseases have been linked to the pathways by which aaRS edits against misacylation involving the non-standard amino acid homocysteine (Hcy).³¹ It is noted that as shown in scheme 1.3 the editing product of Hcy, a cyclic thiolactone, can itself cause protein inactivation via covalent modification of lysyl residues.^{16,32,33}

Scheme 1.3. Schematic illustration of Hcy editing by MetRS followed by attack at, and covalent modification of, a protein's lysyl residue by the resulting thiolactone product.



Hcy misacylation is a potentially devastating mistranslation in protein synthesis. However, there are many other mistranslations that can occur that may also have detrimental consequences. For example, the mis-incorporation of serine into a protein instead of a cysteine would inhibit the protein's ability to form disulfide bridges, a common covalent cross-link involved in protein structure.

This thesis in large part focuses on fundamental aspects of editing mechanisms of aaRS's related to health effects.

1.7 Computational Chemistry

Computational chemistry involves the use of computers to apply the methods of quantum chemistry to predict and study various properties of chemical and biochemical systems. This alternative approach to experimental chemistry allows for detailed examination of high-energy intermediates and transition states. Hence, it is

particularly useful when examining biomolecules such as enzymes. When using this approach on enzymes one can examine and characterize possible mechanisms and the individual roles of catalytically important residues within the active site. The basis for this field is grounded in physics and will be briefly discussed in Chapter 2.

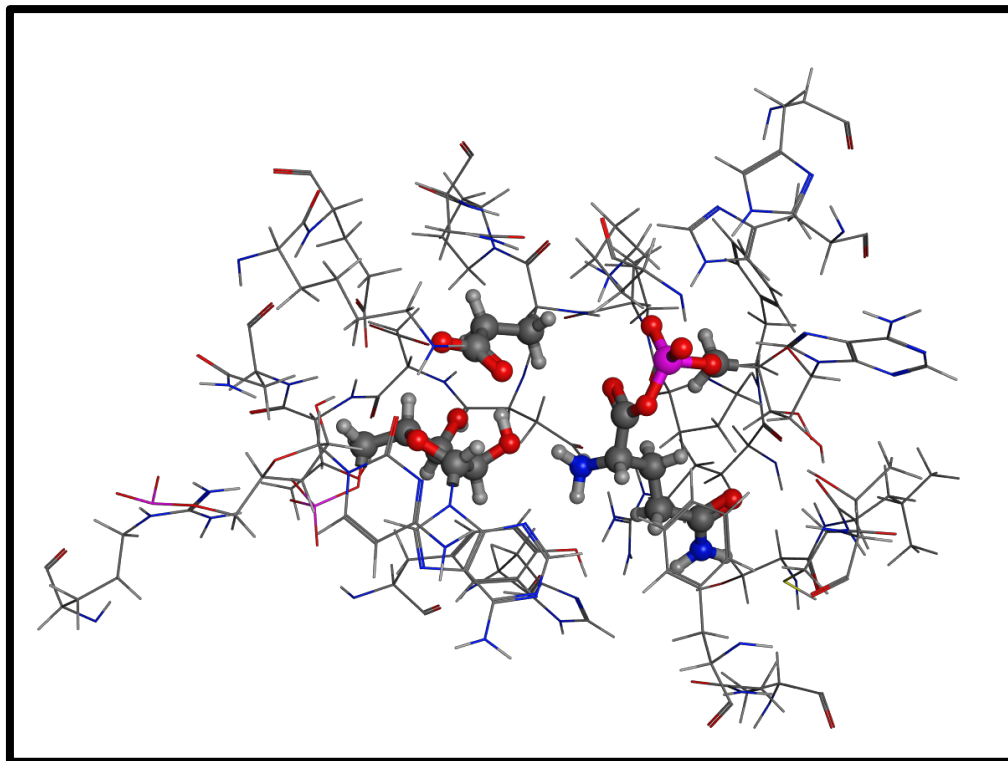
1.8 References

- (1) Martin, W.; Russell, M. J. *Philos. Trans. R. Soc. Lond. Ser. B-Biol. Sci.* **2003**, 358, 59.
- (2) Lehninger, A.; Nelson, D.; Cox, M. *Lehninger Principles of Biochemistry*; W. H. Freeman, 2008.
- (3) Kuo, M. Y. P.; Sharmeen, L.; Dintergottlieb, G.; Taylor, J. J. *Viol.* **1988**, 62, 4439.
- (4) Llano, J.; Gaud, J. W. Mechanistics of Enzyme Catalysis: From Small to Large Active-Site Models. In *Quantum Biochemistry*; Matta, C. F., Ed.; Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim, 2010; Vol. 2.
- (5) Lewis, C. A.; Wolfenden, R. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, 105, 17328.
- (6) Bushnell, E. A. C.; Erdtman, E.; Llano, J.; Eriksson, L. A.; Gaud, J. W. *J. Comput. Chem.* **2011**, 32, 822.
- (7) Zhang, C. M.; Christian, T.; Newberry, K. J.; Perona, J. J.; Hou, Y. M. *J. Mol. Biol.* **2003**, 327, 911.
- (8) Cusack, S. *Curr. Opin. Struct. Biol.* **1997**, 7, 881.
- (9) Lin, L.; Hale, S. P.; Schimmel, P. *Nature* **1996**, 384, 33.
- (10) Helm, M.; Giege, R.; Florentz, C. *Biochemistry* **1999**, 38, 13338.
- (11) Saks, M. E.; Sampson, J. R.; Abelson, J. N. *Science* **1994**, 263, 191.
- (12) Eriani, G.; Delarue, M.; Poch, O.; Gangloff, J.; Moras, D. *Nature* **1990**, 347, 203.
- (13) Hausmann, C. D.; Ibba, M. *FEMS Microbiol. Rev.* **2008**, 32, 705.

- (14) Fournier, G. P.; Andam, C. P.; Alm, E. J.; Gogarten, J. P. *Orig. Life Evol. Biosph.* **2011**, *41*, 621.
- (15) Fournier, G. P.; Andam, C. P.; Alm, E. J.; Gogarten, J. P. *Orig. Life Evol. Biosph.* **2012**, *42*, 377.
- (16) Jakubowski, H. *Acta Biochim. Pol.* **2011**, *58*, 149.
- (17) Hendrickson, T. L.; Schimmel, P. *Translation mechanisms* **2003**, 34.
- (18) Jakubowski, H. *Wiley Interdiscip. Rev.-RNA* **2012**, *3*, 295.
- (19) Jakubowski, H.; Goldman, E. *Microbiol. Rev.* **1992**, *56*, 412.
- (20) Newberry, K. J.; Hou, Y. M.; Perona, J. J. *Embo J.* **2002**, *21*, 2778.
- (21) Bilokapic, S.; Maier, T.; Ahel, D.; Gruic-Sovulj, I.; Soll, D.; Weygand-Durasevic, I.; Ban, N. *EMBO J.* **2006**, *25*, 2498.
- (22) Sankaranarayanan, R.; Dock-Bregeon, A. C.; Rees, B.; Bovee, M.; Caillet, J.; Romby, P.; Francklyn, C. S.; Moras, D. *Nature Struct. Biol.* **2000**, *7*, 461.
- (23) Guth, E.; Connolly, S. H.; Bovee, M.; Francklyn, C. S. *Biochemistry* **2005**, *44*, 3785.
- (24) Huang, W. J.; Bushnell, E. A. C.; Francklyn, C. S.; Gault, J. W. *J. Phys. Chem. A* **2011**, *115*, 13050.
- (25) Minajigi, A.; Francklyn, C. S. *J. Biol. Chem.* **2010**, *285*, 23810.
- (26) Zhang, C. M.; Perona, J. J.; Ryu, K.; Francklyn, C.; Hou, Y. M. *J. Mol. Biol.* **2006**, *361*, 300.
- (27) Pasman, Z.; Robey-Bond, S.; Mirando, A. C.; Smith, G. J.; Lague, A.; Francklyn, C. *S. Biochemistry* **2011**, *50*, 1474.
- (28) Jakubowski, H. *J. Nutr.* **2001**, *131*, 2983S.
- (29) Kim, Y. W.; Kwon, C.; Liu, J. L.; Kim, S. H.; Kim, S. *PLoS One* **2012**, *7*.
- (30) Dewan, V.; Liu, T.; Chen, K. M.; Qian, Z. Q.; Xiao, Y.; Kleiman, L.; Mahasenan, K. V.; Li, C. L.; Matsuo, H.; Pei, D. H.; Musier-Forsyth, K. *ACS Chemical Biology* **2012**, *7*, 761.

- (31) Jakubowski, H. *Phosphorus Sulfur Silicon Relat. Elem.* **2013**, 188, 384.
- (32) Jakubowski, H. *Biochemistry* **1996**, 35, 8252.
- (33) Jakubowski, H. *Cell. Mol. Life Sci.* **2004**, 61, 470.

Chapter 2



Theoretical Methods

2.1 Introduction

Many questions arise when studying chemical systems that can be difficult to examine experimentally. Two particular examples include the chemical reactivity and properties of short-lived intermediates and transition states. Fortunately, there is an alternate approach to elucidate these problems; this is known as computational chemistry. Computational chemistry uses mathematical models provided by classical and quantum mechanics to describe the properties of chemical systems. The chief advantage of using this approach is that unstable or short-lived species can be obtained as easily as long-lived stable species. Since this thesis employs various tools of computational chemistry, here we aim to familiarize the reader with a brief background into some of the methods used as well as the advantages and limitations that also must be considered.

2.2 Classical Mechanics

There are two main fields in physics, classical mechanics and quantum mechanics (discussed later). Classical mechanics concerns itself with the physical laws surrounding the behavior of objects or systems under forces. These laws work well and offer a powerful tool for describing large macroscopic systems but break down at the atomic level. However, they still can provide reasonable approximations for a variety of systems discussed in this chapter.

2.2.1 Molecular Mechanics

Molecular mechanics (MM) takes advantage of the fact that many chemical properties, such as bond stretching, can be described reasonably well using classical mechanics.¹ Instead of trying to describe the motion of electrons, MM concerns itself with the motion of the nuclei. In this approximation, these nuclei are described as charged spheres connected by springs. This approximation uses a number of laws and theorems derived from classical mechanics; its general form is outlined in equation 2.1.¹

$$E^{MM} = \sum E_B + \sum E_A + \sum E_T + \sum E_{vdw} + \sum E_C \quad \text{Eq. 2.1}$$

The total energy in MM is generally described as the sum of several energies, bonds (E_B), angles (E_A) torsion (E_T), Van der Waals (E_{vdw}) and columbic (E_C). The bond energies and angle energies, are described using variations of hooks law to account for changes from the ideal lengths and angles.^{1, 2} E_T is the torsion energy term, which accounts for the torsional strain on the system. The last two terms take into account the non-bonding interactions. Specifically, E_{vdw} is the Vander walls interaction. This is generally taken into account by using the 6-12 Lenard Jones Potential.^{1, 2} Finally, E_c is simply the electrostatic interaction between atoms, which is accounted for by Coulombs law.^{1, 2}

2.2.2 Molecular Dynamics

As described by MM above, Molecular Dynamics (MD) is defined by similar equations. However, where as MM is used to describe a static system, MD simulations

model dynamic systems that change with respect to time. This time dependence is achieved by integrating the Newtonian laws of motion.¹ These equations cannot be solved analytically, but rather the equations are solved for a finite step where the forces are assumed constant. This assumption allows us to solve where the system will be after a finite time interval.¹ The nuclear positions are then updated and the calculation repeats. Once the calculations have been repeated many times, the simulation can be visualized to show the molecular motion with respect to time.

2.2.3 Uses and Limitations

MD and MM have several advantages over the quantum mechanical methods, which will be discussed later. The main advantage of these methods is that they can be computed rapidly. This allows for examination of very large systems such as enzymes or other biomolecules.^{3, 4}

Unfortunately there are several drawbacks to these methods. The main issue is that these methods are empirical in that they need to be parameterized. The parameterization is implemented by studying systems experimentally or with high-level quantum mechanics to determine what the values are for the different terms of eq. 2.1. This means that one should be cautious when using MM and MD on systems that have not been previously parameterized and tested thoroughly. This can, thereby limit the universality of MM and MD calculations. Furthermore, another shortcoming of MM and MD is that they cannot describe bond breaking and bond making to the desired level of accuracy. To do this, quantum mechanics is required.^{1, 4}

2.3 Quantum Mechanics

There are obvious limitations that come with treating molecular systems using classical mechanics that often stem from not correctly describing the nature electrons by effectively ignoring them. Therefore a new theory was implemented and included phenomena such as electron motion and spin. The branch of physics that describes this is referred to as quantum mechanics and was developed in early 20th century by many renowned physicists including Einstein, Plank and Schrödinger, although this list is not exhaustive.

2.4 The Schrödinger Equation

One of the first steps in developing the theory of quantum mechanics started with Erwin Schrödinger and his famous equation aptly named the “Schrödinger Equation” eq. 2.2.⁵

$$\hat{H}\Psi = E\Psi \quad \text{Eq. 2.2}$$

In this equation, \hat{H} is the Hamiltonian operator, E is an eigenvalue, which represents the energy of the system and Ψ is the wave function of the system in question. The wave function is remarkable, as it contains all of the information about the system that one could ever want to know.⁶ Unfortunately, the wave function cannot be directly measured, as it is unobservable. The probability density, $|\Psi|^2$, can however be measured as it represents the electron density.⁶ This brings up one of the fundamental goals of quantum chemistry, which is to try and elucidate Ψ . If we know Ψ and if we

can solve the Hamiltonian we can extract the exact energy of the system. However as we will see later, determining the wave function or energy proves to be exceedingly difficult.

To see the difficulty in calculating the energy, let's first examine the molecular Hamiltonian:

$$\hat{H} = -\sum_{i=1}^{N_E} \frac{1}{2} \nabla^2 - \sum_{a=1}^{N_N} \frac{1}{2M_a} \nabla^2 - \sum_{i=1}^{N_E} \sum_{a=1}^{N_N} \frac{Z_a}{r_{ia}} + \sum_{i=1}^{N_E} \sum_{j>i}^{N_E} \frac{1}{r_{ij}} + \sum_{a=1}^{N_N} \sum_{\beta>a}^{N_N} \frac{Z_a Z_\beta}{r_{a\beta}} \quad \text{Eq. 2.3}$$

Here a and β represent the nuclei and i and j represent the electrons. The first two terms in this equation represent the kinetic energy of the electrons and nuclei respectively. The last three terms represent the columbic attraction of the electrons and nuclei, the repulsions of electrons and repulsions of nuclei respectively.

Due to the complexity of the last three terms, the Schrödinger Equation cannot be solved exactly except in the cases of one-electron one-nuclei systems such as a hydrogen atom. As a result of this inability to compute larger systems, approximations need to be invoked in order to change the Hamiltonian into a form that we are able to deal with for much larger systems.

2.5 Born-Oppenheimer Approximation

The first approximation that will be discussed is the Born-Oppenheimer approximation. This states that since the nuclei are thousands of times heavier than electrons we can assume that the nuclei are essentially fixed with respect to the motion of electrons.⁷ This approximation greatly simplifies the molecular Hamiltonian

since the kinetic energy of the nuclei now goes to zero and the nuclear repulsion term is now reduced to a constant. This resulting Hamiltonian is referred to as the electronic Hamiltonian:⁶

$$\hat{H}_{el} = -\sum_{i=1}^{N_E} \frac{1}{2} \nabla^2 - \sum_{i=1}^{N_E} \sum_{a=1}^{N_N} \frac{Z_a}{r_{ia}} + \sum_{i=1}^{N_E} \sum_{j>i}^{N_E} \frac{1}{r_{ij}} \quad \text{Eq. 2.4}$$

Unfortunately, even though this Hamiltonian is greatly simplified with respect to the original, it is still exceedingly difficult to solve computationally due to the electron-electron repulsion term.⁶

2.6 The Orbital Approximation

Even though Born-Oppenheimer approximation is invoked the resulting electronic Hamiltonian is still too complex to solve. This is because when solving the electronic wave function (ψ_{el}), one must deal with n electrons simultaneously. To simplify this problem the orbital approximation can be invoked. This approximation allows us to treat each electron as an individual spin orbital (χ_i). This is allowed since the orbital approximation assumes that electrons are independent of each other. These spin functions are products of two other functions, the first being a spin function which for electrons are either alpha or beta and a spatial function $\psi_{(x,y,z)}$. While this approach may at first seem inconsequential it greatly simplifies the problem. Now a system of n electrons can be rewritten as a product of n individual spin orbitals:^{4, 6}

$$\psi_{el} = \chi_1(1)\chi_2(2)\chi_3(3) \dots \dots \chi_n(n) \quad \text{Eq. 2.5}$$

The above expression is known as a “Hartree Product” and its simplicity lies in the fact that instead of having to guess at a large complex n-particle wave function, we now only have to determine a series of one electron spin orbitals. This Hartree Product does however have two main shortcomings, the first being that it doesn’t allow for indistinguishability between electrons and second being that the electrons are not antisymmetric with respect to exchange of two electrons.^{4, 6}

These calamities can however be fixed by writing the wave function in the form of a slater determinant:

$$\psi = \frac{1}{\sqrt{n!}} \begin{bmatrix} \chi_1(1) & \cdots & \chi_1(n) \\ \vdots & \ddots & \vdots \\ \chi_n(1) & \cdots & \chi_n(n) \end{bmatrix} \quad \text{Eq. 2.6}$$

This results in a wave function that is antisymmetric with respect to exchange and electrons that are indistinguishable.^{5, 7} The only thing that needs to be done after the orbitals are put in the slater determinant is to normalize the wave function by adding the term in front of the matrix, this ensures the probability of finding an electron over all space is unity.

2.7 Basis Set Expansion

As noted earlier the spin orbital χ_i is composed of a spin function and a spatial function. For molecules or complex chemical systems, these spatial orbitals are molecular orbitals (MO’s). These MO’s can be represented as a linear combination of single electron functions, which are called basis functions:^{1, 4}

$$\psi = \sum_{\mu=1}^N c_{\mu i} \phi_{\mu} \quad \text{Eq. 2.7}$$

In this notation ϕ_μ represents the basis function and $c_{\mu i}$ is the coefficient for the molecular orbitals. In order to get an exact molecular orbital we should have a complete set of basis functions. However this is not common practice since it would require an infinite number of basis functions to form a complete set, so instead only a finite number are used. When the basis functions in Eq. 2.7 are considered to be individual atomic orbitals within a molecule the equation is then known as a linear combination of atomic orbitals (LCAO).^{5, 6}

2.8 The Variational Theorem

The variational theorem states that the energy (E_{Approx}) obtained from any trial wave function ψ_{trial} will be greater than or equal to the exact energy (E_{Exact}) of the real system. This is shown in the following equation:^{4, 7}

$$E_{Approx} = \frac{\int \psi_{trial} \hat{H} \psi_{trial}^* d\tau}{\int \psi_{trial} \psi_{trial}^* d\tau} \geq E_{Exact} \quad \text{Eq. 2.8}$$

This means that if we want to know the exact energy of a system all we have to do is try different ψ_{trial} until we find one that minimizes (E_{Approx}). Fortunately, it has been shown that the energy converges towards the exact energy faster than the trial wave function converges to the exact wave function. This fact has huge implications since it indicates that our guess at a trial wave function doesn't need to be highly accurate to achieve an accurate energy value.

2.9 Hatree-Fock Theory and Wave Function Based Methods

The Hartree-Fock method is one of the most fundamental wave function based methods in computational chemistry. This method uses the central field approximation which, greatly simplifies the notoriously difficult to solve electron-electron repulsion term in the Hamiltonian. The simplification is accomplished by simply integrating repulsion term, resulting in electrons not having specific interactions but instead feeling an average of all other electrons present.^{1, 5}

To solve this new approach a self-consistent field is employed. To understand the meaning behind the term we should first look at the Roothaan-Hall equation:^{4, 6}

$$\sum_{\mu=1}^N C_{\mu i} (F_{\mu\nu} - \varepsilon_i S_{\mu\nu}) = 0 \quad \mu = 1, 2 \dots \dots N \quad \text{Eq. 2.9}$$

where $F_{\mu\nu}$ is the Fock matrix, ε_i is the energy of the molecular orbitals and $S_{\mu\nu}$ is the overlap matrix, which accounts for the overlap between basis functions. If we take into account that the Fock matrix is dependent on the molecular orbital coefficients it becomes apparent how this process is self-consistent. At first an initial guess is made and the Fock matrix will generate a new set of coefficients. These new coefficients will update the Fock matrix which in turn will generate new coefficients, this process will continue until no change occurs thus becoming self consistent. This process is known as the Hartree-Fock self-consistent field procedure.^{4, 6}

Unfortunately this treatment of “electron smearing” can create large problems for even the simplest of models, for example the dissociation of a H₂ molecule and generally underestimating bond lengths.^{4, 6} These failings of Hartree-Fock are due to

the lack of electron correlation. Thus, to improve these results we need to move onto higher electron correlation post Hartree-Fock methods such as configuration interactions and Møller-Plesset Perturbation. By increasing the amount of electron correlation and therefore the accuracy we must greatly increase the required computation time.⁴ While these methods are interesting they will not be discussed in further detail, since they are not used within this thesis.

2.10 Density Functional Theory

Post Hartree-Fock methods have high computation costs. One alternative is to abandon wave function based methods and instead use the modulus square of the wave function or electron density. This approach is called Density Functional Theory (DFT) and has many advantages over wave function based methods. One main advantage is that it can generate results in a time scale that are similar to that of Hartree-Fock, which is the simplest and fastest of all wave function based methods.^{4, 8} Also many DFT-based methods have been shown to have an accuracy and reliability similar to that of more expensive, higher-level wave function-based, electron correlation methods.⁸ Using DFT has another advantage over wave function based methods in that electron density is physically observable unlike the wave function.

2.10.1 Hohenberg-Kohn Theorem

One of the first steps in developing DFT was the establishment of the Hohenberg-Kohn Theorem.^{4, 6, 8} This theorem states that the energy of a given system

is a functional of its electron density. Furthermore, it states that the electron density can be used to determine the systems physical properties.^{4, 6, 8} While this may seem like a minor detail at first, it has large ramifications. In wave function based methods, every electron is a function of four variables: the three spatial and spin. However, with electron density no matter how large the system is, the electron density will only depend on the three spatial coordinates. The Hohenberg-Kohn Theorem while very useful does not specify the functional needed to carry out the calculations.^{4, 6, 8}

2.10.2 Kohn-Sham Theorem

Hohenberg-Kohn Theorem tells us that the electron density is related to the energy, however it does not tell us how to calculate these values. This is when we call upon the Kohn-Sham Theorem. This theorem shows that the exact electron density (ρ) can be described as a sum of single electron density orbitals, ψ_i , which are called Kohn-Sham orbitals.^{4, 6, 8}

$$\rho = \sum_{i=1}^n |\psi_i^{KS}|^2 \quad \text{Eq. 2.10}$$

In addition to showing that the density is a sum of the single electron densities Kohn and Sham also showed that the total energy can be described as the sum of several individual energies.^{4, 8}

$$E_{exact}[\rho] = E^T[\rho] + E^V[\rho] + E^J[\rho] + E^{XC}[\rho] \quad \text{Eq. 2.11}$$

In this expression E^T is the kinetic energy of the electrons in an ideal system, E^V is the potential energy caused by the attraction between electrons and the nucleus in an ideal system, E^J is the energy of repulsion between electrons in an ideal system, and

E^{XC} is the exchange-correlation energy for electrons. This last term includes the corrections for the difference between the real and ideal system. Unfortunately it cannot be solved.^{4, 8} It is common practice however to separate this term into two terms, an exchange term and a correlation term:

$$E^{XC} = E^X + E^C \quad \text{Eq. 2.12}$$

To calculate the exchange and correlation energy a local density approximation (LDA) or generalized-gradient approximation (GGA) is often used. The LDA assumes that the electron density does not vary greatly from point to point, therefore the energy only depends on the electron density at a point and not the gradient. This approach works well for metal solids and other bulk solids, however it doesn't work very well for some molecular systems since the electron density can vary greatly from point to point.^{4, 8} In such cases where electron density changes greatly along the nuclear coordinates, such as polar molecules it is important to take into account the gradient of the electron density.^{4, 8} In particular, DFT functionals such as B3LYP have been used extensively throughout this thesis.

DFT methods are solved using an iterative self-consistent approach that is very similar to the previously described Hartree-Fock method. In particular, an initial set of guessed Kohn-Sham orbitals are used to generate the electron density. This density is then used to generate improved Kohn-Sham orbitals, which then improve the electron density and so on. This process continues until there is little change in the electron density.^{4, 8}

2.11 Basis sets

One of the problems that come about for DFT is the question of how to describe the Kohn-Sham orbitals? The answer to this question is to use a set of basis functions called a basis set to describe the orbitals. These basis functions come in two common types, Slater Type Orbitals (STO) or Gaussian Type Functions (GTF). These two descriptions are very different. The STO's describe the orbitals extremely well but are very computationally expensive, whereas the GTF's do not describe the orbitals very well but are computationally very cheap to use. Luckily, GTF's can be linearly combined to give accurate descriptions of orbitals while remaining much less computationally expensive than STO's.⁴ For this reason GTF's are used more often than STO's.

2.11.1 Split-Valence Basis sets

For chemical reactions and chemical properties it is typically the valence electrons that are of most importance. To take advantage of this fact, computational chemists regularly employ split basis sets. Split basis sets describe each core orbital using one basis function and a minimum of two basis functions per valence orbital. The advantages of using this approach is that you save computational costs on modeling the relatively inert core orbitals yet at the same time gain a better description of the valence orbitals. This inclusion allows for fluctuation in orbital radii providing an overall a better description.⁴ Some common examples of these basis sets are 6-31G (a double zeta basis set) and 6-311G (a triple zeta basis set).

2.11.2 Polarization Basis sets

A problem that arises in computational chemistry is describing the polarization atoms. Take for example a hydrogen atom described by 6-31G; under this description, the hydrogen atom has one symmetric s orbital. However in reality the hydrogen atom may be polarized due to the influence of positive or negative groups around it. Unfortunately the description by 6-31G cannot account for this. However, one way to account for this polarization instead is to add orbitals of higher angular momentum to the calculation.⁴ A common example of this is 6-31G:(d,p) this basis set will add d orbitals to heavy atoms such as carbon and p orbitals to light atoms like hydrogen.

2.11.3 Diffuse Basis sets

The basis sets discussed above can sometimes fail to account for atoms encompassing a larger area, atoms such as anions, or atoms involved in long-range interactions. These effects can often accurately be described through the addition of diffuse functions. Diffuse functions are just large spatial orbitals added to the basis set.⁴ The general notation for diffuse functions is “+” for diffuse functions to be added to heavy atoms and “++” for diffuse functions to be added to light and heavy atoms.

2.12 Solvation

Another issue that can arise when modeling chemical systems is not including environmental effects. In solutions, solvent often plays a role in either stabilizing or

destabilizing chemical compounds. These effects are not taken into account in computational chemistry by default. The environment in which calculations take place is gas phase, which has a dielectric value of one, To properly describe the chemical system, the environment should be taken into account. There are two general ways this can be done. The first is by addition of explicit solvent molecules. This approach however is computationally expensive. Another approach is to implicitly model the solvent.

Implicit solvation is much less computationally expensive and therefore an excellent alternative to modeling the solvent explicitly.⁹ A commonly used approach is the Polarized Continuum Method (PCM). This method specifies contours around the compound onto which a dielectric field is applied.^{4, 10} This dielectric field can be assigned whatever value one wants such as ~80 for water. This approach is often reliable to model general solvation effects but unfortunately it fails to model specific solvent interactions.

2.13 Quantum Mechanics/Molecular Mechanics

In 2013, Warshel and Levitt won a Nobel Prize for their revolutionary approach to study complex chemical systems, Quantum Mechanics/Molecular Mechanics (QM/MM).¹¹ This approach aims to utilize the strengths of quantum mechanics and molecular mechanics synergistically while attempting to minimize the shortcomings of these methods. QM/MM describes the chemically active part of a

system (bonds breaking and making) using quantum mechanics, and the rest of the system using the low cost MM to describe the environment (Figure 2.1).

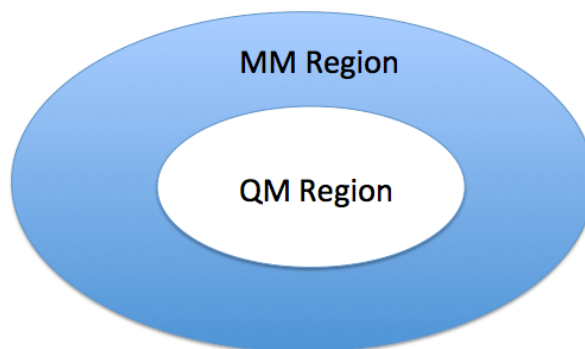


Figure 2.1 An illustration of a QM/MM model in which the environment is modeled at the MM level of theory while the chemically active center uses a QM level of theory.

The advantage with QM/MM is that instead of the couple hundred atoms you are limited to using quantum mechanics, QM/MM allows for the modeling of thousands of atoms.

QM/MM has been found to be a great tool for studying biochemical systems. This is because these systems are often too large to describe using only quantum mechanics, while at the same time the bond making and breaking process is too complex for MM to accurately describe it.¹² Also if truncated models are employed critical residues can be omitted that play vital stabilizing or destabilizing roles in catalysis. If QM/MM is used instead, all of these issues can be addressed, resulting in highly accurate results.¹²

2.14 Potential Energy Surfaces (PES)

The Born-Oppenheimer approximation states that the nuclei are stationary with respect to electrons, due to their substantially larger mass. This results in the energy of a system being a function of the position of the nuclei. One conclusion is that a graph can be plotted with energy on one axis and nuclei position on the other. Such a graph is known as a potential energy surface and an example is shown in Figure 2.2.

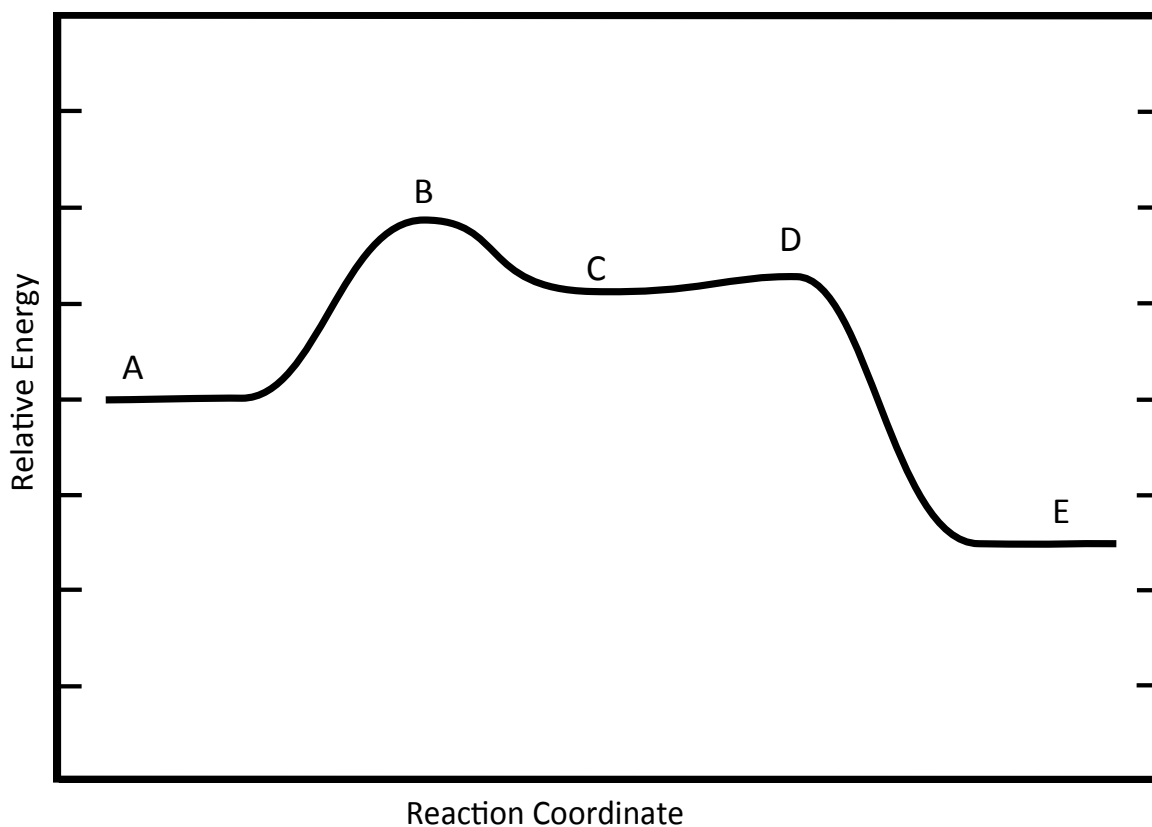


Figure 2.2 A schematic representation of a potential energy surface (PES), where A is the reactant complex, B and D are transition states, C is an intermediate and E is the product complex.¹³

In theory to obtain a complete surface one should compute the energy at every point along the reaction coordinate. Unfortunately, this is rarely feasible due to the

computational cost needed to perform this type of calculation. Moreover, it is not necessary as chemists are often only interested in a few points, namely the maxima and minima of systems. Such points represent the chemically significant reactive complexes, transition states, intermediates and products (see Figure 2.2).

2.15 Technical Aspects and Units

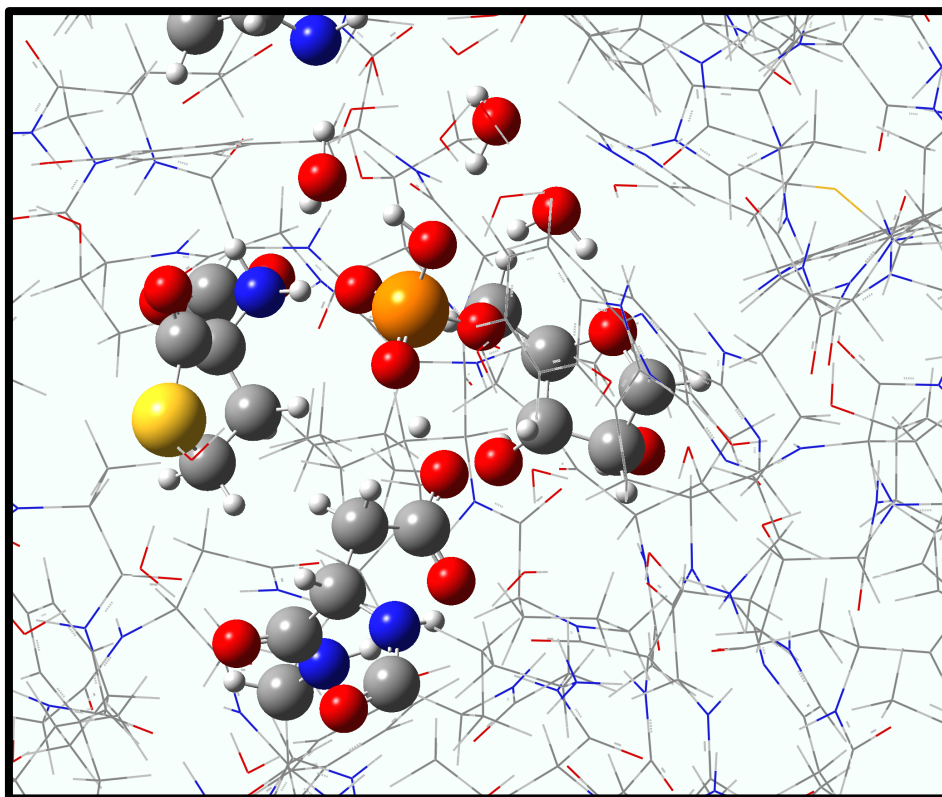
For this computational investigation, all of the calculations were performed using the Gaussian 09 suite.¹⁴ All relative energies reported in this thesis are in kilojoules per mole (kJ mol⁻¹), which were calculated using the defined conversion factor of one Hartree equals 2625.5 kJ mol⁻¹.

2.16 References

- [1] Leach, A. R. (2001) *Molecular Modelling: Principles and Applications* Second ed., Pearson Education Limited:, London.
- [2] Case, D. A., Cheatham, T. E., Darden, T., Gohlke, H., Luo, R., Merz, K. M., Onufriev, A., Simmerling, C., Wang, B., and Woods, R. J. (2005) The Amber biomolecular simulation programs, *J. Comput. Chem.* 26, 1668-1688.
- [3] Bushnell, E. A. C., Huang, W. J., Llano, J., and Gault, J. W. (2012) Molecular Dynamics Investigation into Substrate Binding and Identity of the Catalytic Base in the Mechanism of Threonyl-tRNA Synthetase, *J. Phys. Chem. B* 116, 5205-5212.
- [4] Jensen, F. (1999) *Introduction to Computational Chemistry*, John Wiley & Sons: West Sussex, England.
- [5] Schrödinger, E. (1926) An Undulatory Theory of the Mechanics of Atoms and Molecules, *Physical Review* 28, 1049-1070.

- [6] McQuarrie (2008) *Quantum Chemistry*, 2nd ed., University Science Books, Sausalito, California.
- [7] Born, M., and Oppenheimer, R. (1927) Zur Quantentheorie der Molekeln, *Annalen der Physik* 389, 457-484.
- [8] Koch, W., and Holthausen, M. C. (2001) *A Chemist's Guide to Density Functional Theory*, Wiley - {VCH}, Weinheim - New York.
- [9] Mennucci, B., and Tomasi, J. (1997) Continuum solvation models: A new approach to the problem of solute's charge distribution and cavity boundaries, *J. Chem. Phys.* 106, 5151-5158.
- [10] Tomasi, J., Mennucci, B., and Cammi, R. (2005) Quantum mechanical continuum solvation models, *Chem. Rev.* 105, 2999-3093.
- [11] Warshel, A., and Levitt, M. (1976) THEORETICAL STUDIES OF ENZYMIC REACTIONS - DIELECTRIC, ELECTROSTATIC AND STERIC STABILIZATION OF CARBONIUM-ION IN REACTION OF LYSOZYME, *J. Mol. Biol.* 103, 227-249.
- [12] Sousa, S. F., Fernandes, P. A., and Ramos, M. J. (2012) Computational enzymatic catalysis - clarifying enzymatic mechanisms with the help of computers, *Phys. Chem. Chem. Phys.* 14, 12431-12441.
- [13] Bushnell, E. A. C., Huang, W., and Gault, J. W. (2012) Applications of Potential Energy Surfaces in the Study of Enzymatic Reactions, *Advances in Physical Chemistry* 2012, 15.
- [14] Frisch, M. J.; et al. Gaussian 09, revision D.02; Gaussian, Inc.: Wallingford, CT, 2009

Chapter 3



**Enzymatic and Substrate-Assisted
Editing of Highly Reactive
Nonstandard Amino Acids by
Aminoacyl-tRNA Synthetases**

3.1 Introduction

Aminoacyl-tRNA synthetases (AaRS's), in addition to their central role in protein biosynthesis, are important for a diverse range of biochemical processes including inflammation, cell death and viral assembly.¹⁻³ In particular, they catalyze the activation of their corresponding amino acid and the subsequent aminoacylation of their cognate tRNA.^{4,5} Consequently they have the difficult yet essential task of discriminating between amino acids. This can be particularly problematic for aaRS's whose cognate amino acid is structurally similar or isoelectronic with one another. This is illustrated well in ThrRS where it must discern between threonine, serine and valine.⁶⁻⁸ This matter is further complicated since aaRS's must also discriminate against other species, for example, non-standard amino acids such as homocysteine (Hcy) and homosereine (Hse).⁹⁻¹¹ Indeed, the latter are highly reactive and their incorporation into proteins has been implicated in a number of diseases including stroke, cancer and Alzheimer's.¹²

In response, several aaRS's are known to utilize a second active site whose function is to remove, or edit the aminoacyl moiety of misacylated tRNA. For some aaRS's however it is thought that the aminoacylation's active site itself may be involved in editing. For instance, MetRS is able to discriminate and edit against Hcy. Based on experimental studies, it has been proposed that the MetRS aminoacylation active site has a thiol specific subsite, as illustrated in Figure 3.1.^{10,11,13,14} Consequently, unlike the native substrate Met, Hcy preferentially binds such that its thiol is positioned near its carboxyl carbon and thus self-cyclization, i.e. editing, is enhanced.

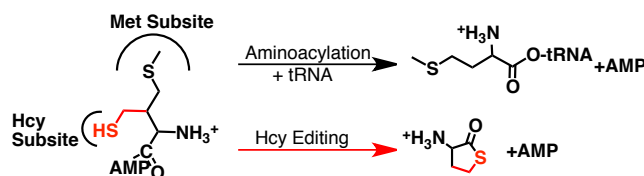


Figure 3.1. The proposed editing mechanism of Hcy in MetRS by which a thiol subsite facilitates self-cyclization.^{10,11,15}

However, this proposal is not without some controversy.¹⁶ In particular, site-directed mutagenesis studies were unable to definitively confirm or refute the proposed thiol subsite.¹⁶ Furthermore, it has been noted that Hcy is arguably the most reactive amino acid. Hence, simply by releasing Hcy-AMP from the active site of MetRS active site, self-cyclization may occur.¹⁶

3.2 Theoretical Methods

Herein a complementary computational approach of molecular dynamics (MD) and quantum mechanical/molecular mechanics (QM/MM) methods has been used to investigate the mechanism by which MetRS edits against Hcy. A detailed description of the computational protocol used is given in the Supporting Information (Chapter 4). The Molecular Operating Environment (MOE) program was used for all molecular dynamics (MD) simulations.¹⁷ All QM/MM calculations were performed using the ONIOM formulism within the Gaussian 09 program.¹⁸⁻²⁶ Optimized structures, frequencies and Gibb's Free Energy corrections (ΔE_{Gibbs}) were obtained at the ONIOM(B3LYP/6-31G(d,p):AMBER96) level of theory within the mechanical embedding (ME) formalism.²⁷⁻³² Relative free energies were obtained via single points on these structures at the ONIOM(B3LYP/6-311G(2df,p):AMBER96)-ME level of theory with inclusion of the appropriate Gibb's correction.²⁷⁻³²

3.3 Results and Discussion

Initially, the binding within the active site of MetRS of the native substrate, Met-AMP, as well as the possible alternate substrates Hcy-AMP and Hse-AMP were examined. An X-Ray crystal structure (PDB ID: 2CT8)³³ of MetRS with bound Met-AMP substrate analog 5'-O-[(L-methionyl)-sulfamoyl]adenosine was modified accordingly (Chapter 4). Average structures of the 3 bound substrates obtained after 10 ns MD simulations are overlaid in Figure 3.2. Remarkably, as can be seen they all preferentially bind in the same "linear" position within the active site. Indeed, while the largest differences are observed between the tails of their aminoacyl R-group's, each remains quite similarly oriented.

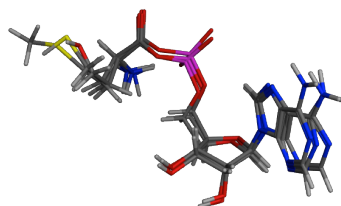


Figure 3.2. Overlaid average MetRS active site-bound structures of Met-AMP, Hcy-AMP and Hse-AMP.

In order to investigate possible mechanisms by which Hcy-AMP may cyclize within the active site of MetRS, a suitable chemical model for use in conjunction with an ONIOM(QM/MM) approach was derived from the average MD structure of the MetRS...Hcy-AMP complex (Chapter 4). For the cyclization reaction to occur, a base must facilitate deprotonation of the thiol of the Hcy moiety in order for its sulfur to nucleophilically attack Hcy's sp^2 carbonyl carbon (C_{carb}). Analogous to that generally proposed to occur in aaRS's, the ability of the Hcy-AMP's own phosphate to act as the required base was examined, i.e., a substrate-assisted editing mechanism.^{34,35} Unfortunately, the average distance between Hcy's sulfur and the nearest phosphate

oxygen in the initial reactive complex (RC) is 7.0 Å (Chapter 4). Hence, the substrate must undergo a large conformational change in order to react. This step, going from a linear to bent conformation occurs via **TS1** with a barrier of 27.5 kJ mol⁻¹ to give the alternate substrate-bound complex **I1** (3.3a). This is followed by nucleophilic attack of Hcy's sulfur at C_{carb} with concomitant intramolecular proton transfer between the thiol and phosphate of Hcy. This occurs with a significantly larger barrier of 98.2 kJ mol⁻¹ via the tetrahedral cyclic transition state **TS2**, and leads directly to formation of the product complex (**PC**), i.e., bound AMP and a cyclic Hcy derivative. This is in fact the rate-limiting step. Importantly, it is comparable to the calculated rate-limiting step barriers for aminoacylation in some aaRS's.^{6,36} Thus, mis-aminoacylation by MetRS of its cognate tRNA with Hcy appears to be competitive with its editing of Hcy.

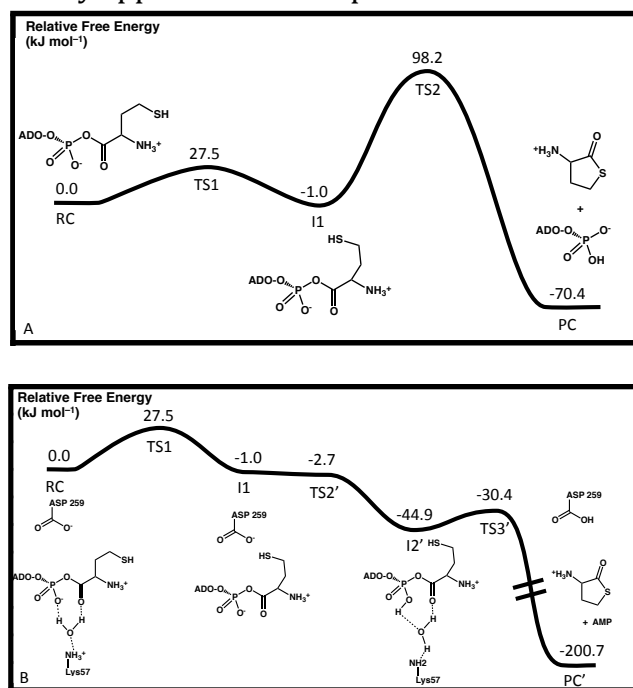


Figure 3.3. Free Energy surfaces (kJ mol⁻¹) for "editing" of Hcy-AMP in MetRS using as the mechanistic base (A) a Hcy-AMP phosphate non-bridging oxygen, or (B) Asp259.

No other potential bases within the substrate, e.g., the aminoacyl's amine group as recently suggested to be the base within ThrRS, appear to be suitably positioned.⁶ Examination of the substrate's required conformational change and the surrounding enzyme residues, however, suggests a possible alternate enzyme-based candidate, Asp259. Notably, it is much closer to the thiol of Hcy in the initial **RC** with an $\text{Asp259COO}^- \cdots (\text{H})\text{S}_{\text{Hcy}}$ distance of 5.30 Å. Concomitantly, the average initial distance between the carboxyl of the Asp259 and the substrates C_{carb} center is shorter at 4.25 Å. One apparent advantage of using Asp259 as a base instead of the phosphate is that instead of having to undergo large conformational changes, only a dihedral rotation about Hcy's side-chain $\text{C}_\beta\text{--C}_\gamma$ bond is required (Figure 3.3). Indeed, this step occurs via **TS1** with a barrier of only 27.5 kJ mol⁻¹. In the intermediate formed, **I1**, the thiol of Hcy is now well positioned to both transfer its proton to Asp259 and nucleophilically attack C_{carb} . In particular, the $\text{HcyS} \cdots \text{C}_{\text{carb}}$ distance has decreased from 4.19 to 3.29 Å, while the $\text{Asp259COO}^- \cdots (\text{H})\text{S}_{\text{Hcy}}$ distance has now shortened dramatically by almost 2 Å to 3.31 Å. It is noted that the $\text{Asp259COO}^- \cdots \text{C}_{\text{carb}}$ distance has increased from 4.25 to 4.95 Å (Chapter 4). Importantly, this step is the overall rate-limiting step of this alternate mechanism.

Subsequently, a barrierless proton transfer via **TS2'** occurs from the R-group amine of the active site residue Lys57 through a water onto the substrates phosphate. The resulting intermediate **I2'** lies 44.9 kJ mol⁻¹ lower in energy than **RC**. The final step is transfer of the thiol proton of Hcy to the carboxyl of Asp259 with concomitant attack of the Hcy sulfur at C_{carb} . This occurs via **TS3'** with a barrier of just 14.5 kJ mol⁻¹ to give the product complex (**PC**); bound cyclic thiolactone and AMP. This step is analogous to the initial pathway in which the substrates phosphate acted as the base, but occurs with a barrier that is a factor of ~7 lower.

However, as has been suggested previously Hcy is arguably one of the most reactive amino acids that could be incorporated into a protein.¹⁵ This is due in part to the pK_a of its thiol and thus it is possible that Hcy-AMP may also non-enzymatically self-cyclize in solution. Its oxygen analogue Hse, however, is less reactive and its R-group hydroxyl has a higher pK_a . Thus, we also examined the ability of Asp259 to act as a suitable mechanistic base to allow MetRS to edit Hse. As for MetRS...Hcy-AMP, an appropriate chemical model for use in conjunction with ONIOM(QM/MM) was derived from the average MD structure (see above). It is noted that in the initial reactant complex the $_{Asp259}COO^- \cdots C_{carb}$ distance is slightly larger than observed for Hcy-AMP at 4.45 Å. However, the barrier for rotation about the C_b-C_g bond of Hse-AMP within the active site of MetRS, is lower than that obtained for Hcy-AMP at only 19.1 kJ mol⁻¹ above the initial reactant complex to give an intermediate lying lower in energy than **RC** by 39.1 kJ mol⁻¹ (see Chapter 4). Upon rotation the $_{Asp259}COO^- \cdots (H)O_{Hse}$ and $C_{carb} \cdots (H)O_{Hse}$ distances have decreased from 4.76 and 3.59 Å to 2.81 and 3.03 Å, respectively. Meanwhile the $_{Asp259}COO^- \cdots C_{carb}$ has decreased from 5.00 to 4.73 Å. The remainder of the mechanism is similar to that obtained for Hcy-AMP, except that the barriers are now slightly higher. More specifically, the subsequent step is a proton transfer from the side chain amine of Lys57 to the forming oxyanion of Hse-AMP along with the attack of Hse's oxygen and the deprotonation of the hydroxyl group by Asp259. This proceeds with a barrier of 11.1 kJ mol⁻¹ and gives a tetrahedral intermediate lying 75.8 kJ mol⁻¹ lower in energy than **RC**. The final step is concomitant proton transfer from the newly formed alcohol to the phosphate of Hse-AMP and the cleavage of the $PO-C_{carb}$ bond to give AMP and a cyclic lactone Hse derivative. Notably, this reaction step has a slightly higher barrier than obtained for Hcy-AMP of 30.9 kJ mol⁻¹ and in fact is the rate-limiting step in editing of Hse-AMP by MetRS. It is also

noted that this reaction step is similar to the second-half reaction of aaRS's; aminoacylation of their cognate tRNA via deprotonation of an alcohol with concomitant nucleophilic attack at an sp^2 carbon, but now, has a significantly lower barrier.^{6,36}

Thus for both Hcy-AMP and Hse-AMP the active site residue Asp259 appears able to effectively act as a mechanistic base in editing by MetRS. Moreover, it has been noted that other aaRS's must also edit against Hcy and Hse. For example, ValRS, IleRS and LeuRS all bind these nonstandard amino acids.⁹ Thus, we examined whether their active sites may also contain a similarly positioned and suitable residue.

Both Hcy-AMP and Hse-AMP were mutated from the substrate crystalized within the active sites of ValRS (PDB ID: 1GAX), IleRS (PDB ID: 1WK8) and LeuRS (PDB ID: 3ZGZ), followed by 10 ns MD simulations (Chapter 4).³⁷⁻³⁹ For ValRS and LeuRS both substrates lie in the same approximate position as observed in MetRS. Furthermore, both enzymes contain an Asp (ValRS) or Glu (LeuRS) residue in the same approximate position as Asp259 in MetRS. In order to quantify whether they may also potentially participate in editing, the average distance between the nearest carboxyl oxygen of the Asp or Glu residue and the substrate's C_{carb} center was determined and is given in Table 3.1.

For LeuRS the average $Glu_{532}COO^- \cdots C_{carb}$ distances for both Hcy-AMP and Hse-AMP are comparable to those observed for the same substrate bound in MetRS. In the case of ValRS the nearest carboxyl oxygen of the acidic active site residue Asp490 is further away from C_{carb} by on average 1.01 and 1.39 Å for Hcy-AMP and Hse-AMP respectively (see Table 3.1). However, this does not necessarily preclude Asp490 from acting as the editing base; in the case of ValRS...Hse-AMP MD simulations suggest that a water molecule can sit between $Asp_{490}COO^-$ and C_{carb} .

Table 3.1. Average distances (Angstroms) between C_{carb} and the nearest Asp/Glu-COO⁻ oxygen from MD simulations (see text).

| AaRS / Residue | $r(\text{COO}^- \cdots \text{C}_{\text{carb}})$ | |
|-------------------|---|-------------------|
| | Hcy-AMP | Hse-AMP |
| MetRS / Asp259 | 4.25 | 4.45 ^a |
| LeuRS / Glu532 | 4.39 | 4.31 |
| ValRS / Asp490 | 5.26 | 5.84 |

^a Average taken over 5ns.

IleRS does not have an Asp or Glu residue in its active site that lies near C_{carb} of Hcy- or Hse-AMP. Intriguingly, however, the conformation of these substrates when bound in IleRS is markedly different from that observed for MetRS, LeuRS and ValRS (Figure 3.4). In particular, Hcy-AMP binds within IleRS in a "bent" conformation. As a result, its own phosphate is now positioned such that it may be able to act as the mechanistic base (Figure 3.4). That is, IleRS, which lacks a potential base in its active site may use substrate-assisted editing. Indeed, in IleRS...Hcy-AMP the shortest average intramolecular distance between the phosphate and thiol of Hcy-AMP is 4.55 Å, 2.44 Å shorter than obtained for MetRS...Hcy-AMP.

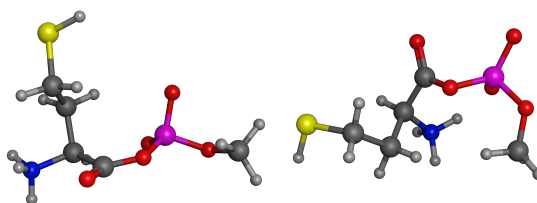


Figure 3.4. An illustration of the conformation of Hcy-AMP when bound in the active site of IleRS (left) and MetRS (right).

3.4 Conclusions

The present results suggest that the MetRS active site residue Asp259 can play an essential role as a mechanistic base in editing against the non-standard amino acids Hcy and Hse. That is, MetRS can exploit enzymatic editing for such amino acids. The rate-limiting step for Hcy-AMP cleavage via intramolecular cyclization in MetRS, is the initial required rotation about the substrates aminoacyl C_b-C_g bond with a barrier of just 27.5 kJ mol⁻¹. In addition, both ValRS and LeuRS, which likely also edit against Hcy and Hse, have similar active site motifs to MetRS. In particular, all contain a similarly positioned residue carboxylate be it either an Asp (MetRS and ValRS) or Glu (LeuRS). Furthermore, Hcy-AMP and Hse-AMP preferentially bind in similar "linear" positions in all 3 aaRS's: MetRS, LeuRS and ValRS. In contrast, IleRS does not contain an acidic active site residue but Hcy-AMP and Hse-AMP instead bind in a "non-linear" position. As a consequence, the non-bridging oxygen's of the substrates own phosphate group are now better positioned to facilitate intramolecular nucleophilic attack of the Hcy- and Hse-AMP's thiol or hydroxyl group, respectively.

3.5 References

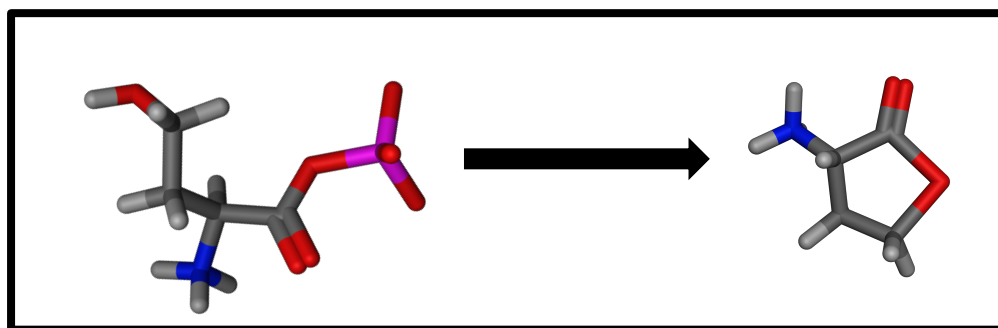
- (1) Hausmann, C. D.; Ibba, M. *FEMS Microbiol. Rev.* **2008**, 32, 705.

- (2) Dewan, V.; Liu, T.; Chen, K. M.; Qian, Z. Q.; Xiao, Y.; Kleiman, L.; Mahasenan, K. V.; Li, C. L.; Matsuo, H.; Pei, D. H.; Musier-Forsyth, K. *ACS Chemical Biology* **2012**, 7, 761.
- (3) Wang, J.; Fang, P. F.; Schimmel, P.; Guo, M. *J. Phys. Chem. B* **2012**, 116, 6798.
- (4) Cusack, S. *Curr. Opin. Struct. Biol.* **1997**, 7, 881.
- (5) Safro, M. G.; Moor, N. A. *Mol. Biol.* **2009**, 43, 211.
- (6) Huang, W. J.; Bushnell, E. A. C.; Francklyn, C. S.; Gault, J. W. *J. Phys. Chem. A* **2011**, 115, 13050.
- (7) Minajigi, A.; Francklyn, C. S. *J. Biol. Chem.* **2010**, 285, 23810.
- (8) Sankaranarayanan, R.; Dock-Bregeon, A. C.; Rees, B.; Bovee, M.; Caillet, J.; Romby, P.; Francklyn, C. S.; Moras, D. *Nat. Struct. Biol.* **2000**, 7, 461.
- (9) Jakubowski, H. *J. Nutr.* **2001**, 131, 2983S.
- (10) Jakubowski, H. *Acta Biochim. Pol.* **2011**, 58, 149.
- (11) Jakubowski, H. *WIREs* **2012**, 3, 295.
- (12) Jakubowski, H. *Cell. Mol. Life Sci.* **2004**, 61, 470.
- (13) Jakubowski, H.; Fersht, A. R. *Nucleic Acids Res.* **1981**, 9, 3105.
- (14) Jakubowski, H. *Biochemistry* **1996**, 35, 8252.
- (15) Jakubowski, H. *Phosphorus Sulfur Silicon Relat. Elem.* **2013**, 188, 384.
- (16) Hendrickson TL, Schimmel P. Transfer RNA-Dependent Editing by Aminoacyl-tRNA Synthetases, in Translation Mechanisms. In: Lapointe J, editor. Kluwer Academic Press; 2003. pp. 34–64.
- (17) *Molecular Operating Environment (MOE)*; Chemical Computing Group Inc: 010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7,, 2012.
- (18) Frisch, M. J.; et al. Gaussian 09, revision D.02; Gaussian, Inc.: Wallingford, CT, 2009

- (19) Bearpark, M. J.; Ogliaro, F.; Vreven, T.; Boggio-Pasqua, M.; Frisch, M. J.; Larkin, S. M.; Morrison, M.; Robb, M. A. *J. Photochem. Photobiol., A* **2007**, *190*, 207.
- (20) Dapprich, S.; Komaromi, I.; Byun, K. S.; Morokuma, K.; Frisch, M. J. *Theochem*. **1999**, *461*, 1.
- (21) Humbel, S.; Sieber, S.; Morokuma, K. *J. Chem. Phys.* **1996**, *105*, 1959.
- (22) Maseras, F.; Morokuma, K. *J. Comput. Chem.* **1995**, *16*, 1170.
- (23) Svensson, M.; Humbel, S.; Froese, R. D. J.; Matsubara, T.; Sieber, S.; Morokuma, K. *J. Phys. Chem.* **1996**, *100*, 19357.
- (24) Vreven, T.; Byun, K. S.; Komaromi, I.; Dapprich, S.; Montgomery, J. A.; Morokuma, K.; Frisch, M. J. *J. Chem. Theory Comput.* **2006**, *2*, 815.
- (25) Vreven, T.; Morokuma, K. *J. Comput. Chem.* **2000**, *21*, 1419.
- (26) Vreven, T.; Morokuma, K.; Farkas, O.; Schlegel, H. B.; Frisch, M. J. *J. Comput. Chem.* **2003**, *24*, 760.
- (27) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.
- (28) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 1372.
- (29) Lee, C. T.; Yang, W. T.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.
- (30) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. *J. Comput. Chem.* **2005**, *26*, 1668.
- (31) Handy, N. C.; Cohen, A. J. *Mol. Phys.* **2001**, *99*, 403.
- (32) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623.
- (33) Nakanishi, K.; Ogiso, Y.; Nakama, T.; Fukai, S.; Nureki, O. *Nat. Struct. Mol. Biol.* **2005**, *12*, 931.
- (34) Banik, S. D.; Nandi, N. *J. Biomol. Struct. Dyn.* **2012**, *30*, 701.

- (35) Perona, J.; Gruic-Sovulj, I. Synthetic and Editing Mechanisms of Aminoacyl-tRNA Synthetases, In *Aminoacyl-tRNA Synthetases in Biology and Medicine*; Kim, S., Ed.; Springer Netherlands: 2014; Vol. 344, pp. 1-41.
- (36) Liu, H. N.; Gault, J. W. *J. Phys. Chem. B* **2008**, *112*, 16874.
- (37) Fukai, S.; Nureki, O.; Sekine, S.; Shimada, A.; Tao, J. S.; Vassylyev, D. G.; Yokoyama, S. *Cell* **2000**, *103*, 793.
- (38) Fukunaga, R.; Yokoyama, S. *J. Mol. Biol.* **2006**, *359*, 901.
- (39) Chopra, S.; Palencia, A.; Virus, C.; Tripathy, A.; Temple, B. R.; Velazquez-Campoy, A.; Cusack, S.; Reader, J. S. *Nat. Commun.* **2013**, *4*.

Chapter 4



Mechanistic Insights into the Self-Cyclization of Homocysteine and Homoserine

4.1 Introduction

The editing of aaRSs is of the upmost importance for the proliferation of all living things is directly dependent on their correct functioning. Occurrences of incorrect editing may result in many illnesses.¹ MetRS in particular has to edit against non-standard homocysteine (Hcy).² In this case, if this important editing cannot occur, cardiovascular disease along with other diseases may result.³ The aim of this chapter is to provide greater mechanistic details along with structural information to complement the results of chapter 3 to provide a more complete picture of the process of editing Hcy.

4.2 Theoretical methods

To obtain a structure that would best describe the native enzyme, a protocol was carried out as described herein. First, all molecular dynamics simulations were carried out using the Molecular Operating Environment (MOE) software.⁴ Second, all crystal structures had their solvent waters deleted then were resolvated and simulated for 10 ns. Finally, structures were then clustered into 10 groups based on the root mean square deviation (RMSD's) of the active site and the average structure of the most populated cluster was chosen for the QM/MM model. This model was set up to include the substrate and potential catalytic residues along with any active site residues that could provide important electrostatic interactions in the high layer. The atoms within the high layer were allowed to be mobile along with all amino acid in the low layer that lies within 10 Å from each atom of the substrate, the remaining amino

acids were fixed to allow for the overall structure of the enzyme to remain unperturbed. QM/MM calculations were performed using the ONIOM formalism within the Gaussian 09 program suite.⁵⁻¹³ Optimized structures were calculated at the ONIOM (B3LYP 6-31G(d,p):AMBER96) level of theory followed by single points at ONIOM (B3LYP/6-311G(2df,p):AMBER96//B3LYP/6-31G(d,p):AMBER96) + ΔE_{Gibbs} level of theory within a mechanical embedding (ME) formalism.¹⁴⁻¹⁹ These energies were used to construct potential energy surfaces to represent the thermodynamics of catalysis. Values on potential energy surfaces with an * indicate that the transition state was found by a scan rather than a TS optimization.

4.3 Phosphate facilitated cyclization in MetRS

The prototypical base in the transfer step of aaRS's is the phosphate group of AMP.²⁰ Therefore, having the phosphate act as a base in the cyclization of Hcy was the first mechanism tested. After the molecular dynamics results were interpreted, it was observed that the distance between the phosphate's non-bridging oxygen and the thiol sulfur was long, on average 6.99 Å (figure 4.1).

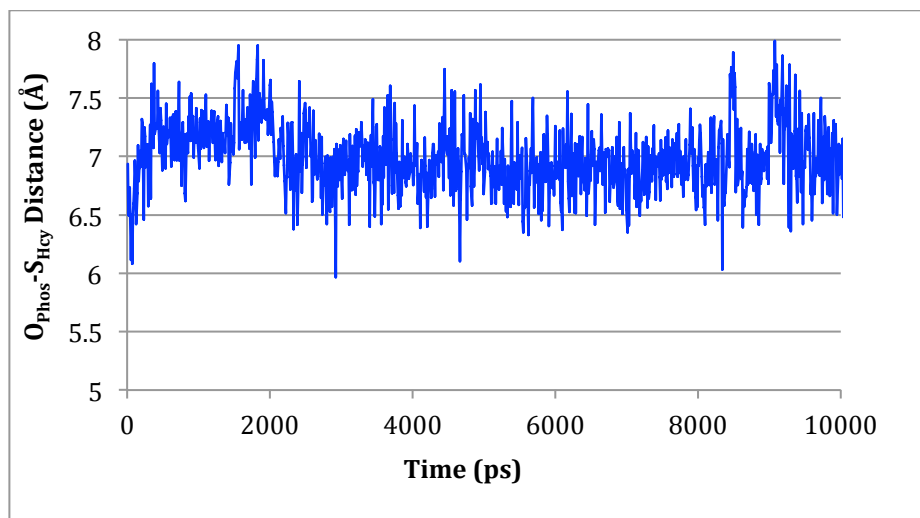


Figure 4.1. Molecular dynamics simulation of MetRS showing the distance between the phosphates non bridging oxygen and Hcy's sulfur with respect to time.

The distance between the oxygen and sulfur potentially has direct implications on the energetics of the cyclization of Hcy. In the present case, the large distance observed still allows for cyclization to occur. The rate-limiting step for cyclization was a transition state that has a sulfur hydrogen distance of 1.78 Å and an oxygen hydrogen distance of 1.12 Å (figure 4.2). The sulfur carbon distance on the other hand is 2.60 Å. Furthermore, the barrier for this reaction is 98.2 kJ mol⁻¹. Although this barrier is enzymatically feasible, it is comparable to the rate-limiting step in the transfer of aa's to tRNA by aaRS's.^{21,22} This suggests that the editing of Hcy should be kinetically competitive with the transfer of aa's.

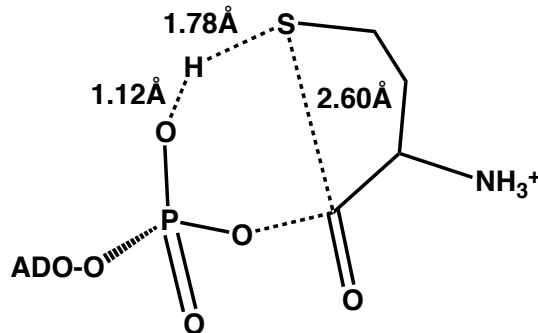


Figure 4.2. A schematic representation of the distances calculated in the phosphate base mediated self-cyclization of Hcy.

Interestingly, this may imply that the most energetically costly part of this reaction is the proton abstraction. This high barrier is not chemically intuitive since the editing reactions should proceed quicker than the acylation reaction due to the fact that both the editing and transfer take place in the same active site. Furthermore, there is no incorporation of Hcy onto tRNA due to the effectiveness of the editing site of Hcy.^{23,24} Therefore, the high barrier for Hcy cyclization using phosphate as a base suggests that AMP's phosphate may not be the correct base for this reaction.

4.4 An alternative the base in the self-cyclization of Hcy

To look for an alternative base we examined the binding orientation of Hcy-AMP within the active site. This has been proposed to bind in a different orientation than the natural substrate, Met-Amp. However, when sequentially comparing the positions of the α -carbons in the active site of MetRS when complexed with Met-AMP and then Hcy-AMP, we observed a RMSD of 0.38 Å. This is indicative of very similar

natural binding orientations. Thus, the same alternative base would likely do the job regardless of the substrate. After revisiting the crystal structure (2CT8), Asp259 was found to be a good candidate as it was located 5.48 Å from the sp^2 carbon.²⁵ A molecular dynamic simulation was performed to get an average structure to obtain a more representative depiction of the native bond distance the oxygen and the carbonyl carbon (figure 4.3) and an average distance of 4.25 Å was observed. However, close proximity to the carbonyl carbon itself is not enough for Asp259 to act as a base; it also needs to be close to Hcy's thiol from which it will abstract a proton. Thus, the Asp259- S_{Hcy} distance was monitored (figure 4.4) and an average of 5.30 Å was calculated.

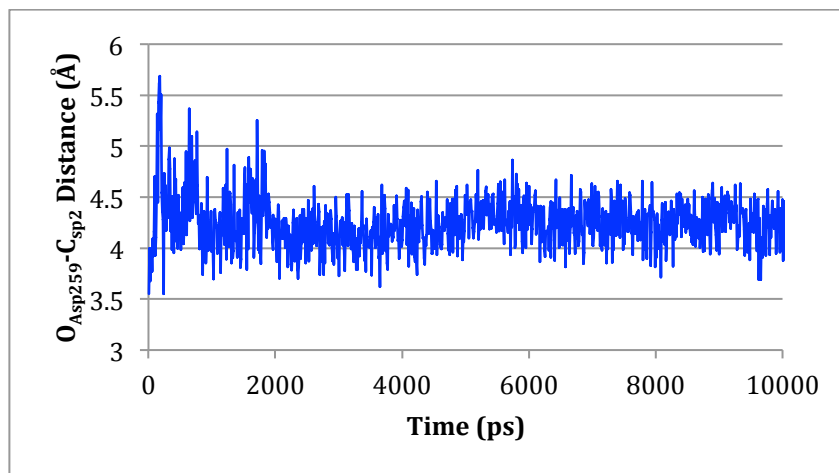


Figure 4.3. A graph showing the distance between the Asp259 and the Hcy's carbonyl carbon with respect to time during the molecular dynamics simulation of MetRS.

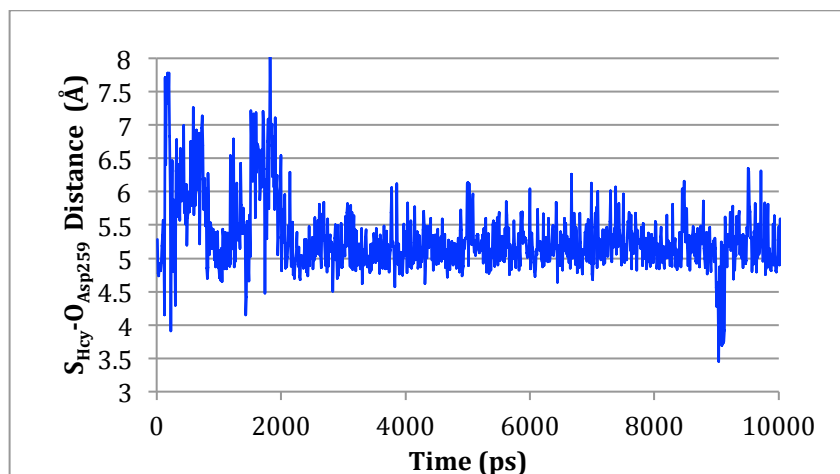


Figure 4.4. A graph showing the distance between the Asp259 and the Hcy's sulfur with respect to time during the molecular dynamics simulation of MetRS.

The average distance of 5.30 Å is much less than the phosphate sulfur distance of 6.99 Å, which could indicate that Asp259 is more likely to act as a base during the cyclization of Hcy. Therefore, to investigate this we employed QM/MM calculations.

4. 5 Mechanism of Hcy cyclization

The editing reaction of Hcy starts with a reactive complex where the sulfur of Hcy is 5.20 Å away from Asp259's oxygen and 3.94 Å away from the carbonyl carbon (scheme 4.1). The 5.20 Å distance between the thiol and carboxylate is large, however after a dihedral rotation for which the barrier is 27.5 kJ mol⁻¹ (figure 4.5) the thiol was found to be 3.31 Å from the carboxylate and 3.29 Å from the sp² carbon.

Scheme 4.1. An illustration a reaction pathway for the cyclization of Hcy in MetRS utilizing Asp259 as a mechanistic base.

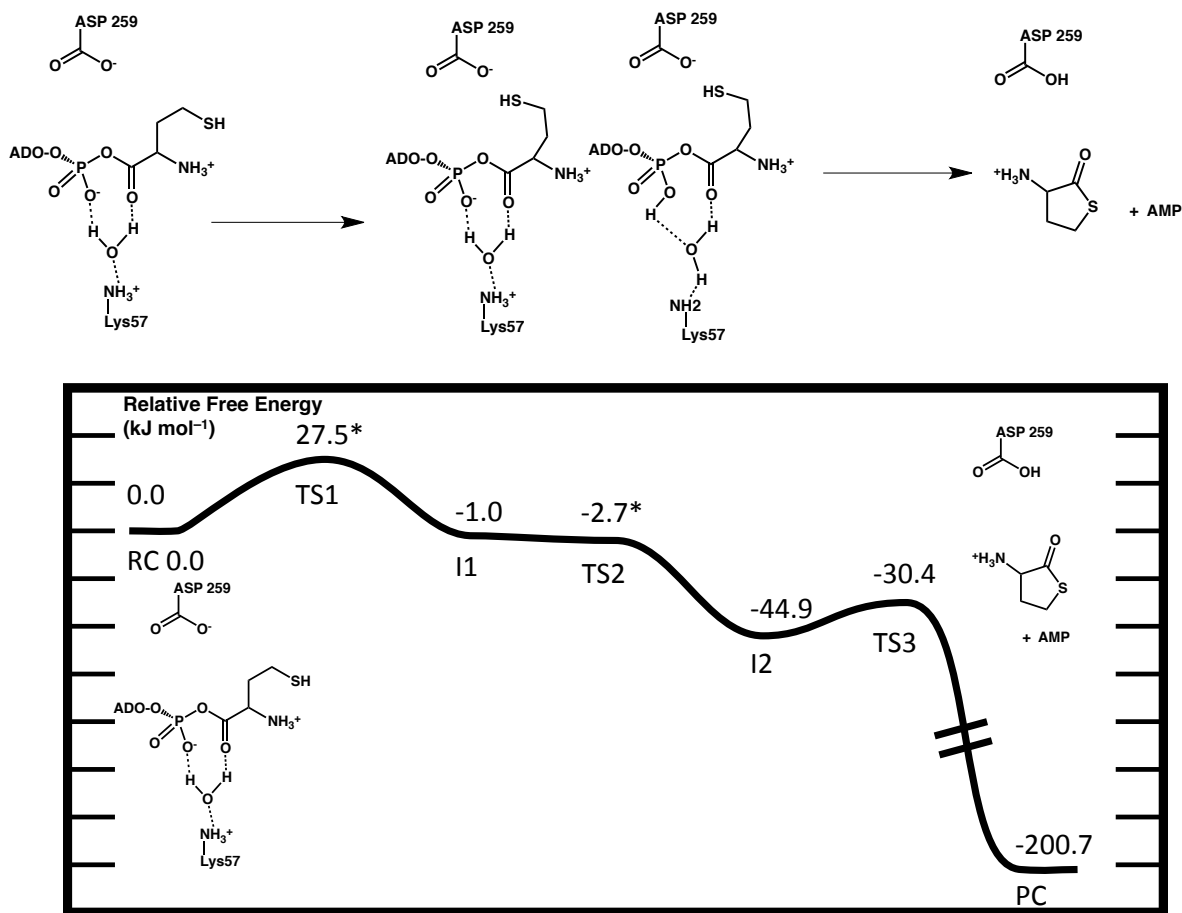


Figure 4.5. The potential energy surface for the cyclization reaction of Hcy-AMP in MetRS with all relative energies reported in kJ mol⁻¹.

Hcy is now favourably positioned to attack the carbonyl carbon and have its proton abstracted. This is indicated by a 1.95 Å hydrogen bond between the thiol's hydrogen and Asp259. However, before this attack can occur Lys57 must transfer a proton to the phosphate to make it a better leaving group. This is a probable

consequence of gas phase calculations that stabilize neutral charges and destabilize charged species. This step has no barrier when free energy corrections are added and results in a more favourable complex lying 44.9 kJ mol^{-1} lower in energy than the reactive complex. After the proton shuttle, the thiol was found to be 3.24 \AA from the carbonyl carbon.

Once the phosphate is protonated the thiol can finally attack the carbonyl carbon. As the thiol moves towards the carbon it loses its proton to Asp259. The energy continues to rise until it reaches the transition state, where the sulfur oxygen distance is 2.04 \AA and the $\text{O}_{\text{Asp259}}\text{-H}_{\text{Hcy}}$ distance is 0.99 \AA . This step is concerted similar to the mechanism where the phosphate is the base and the thiol attacks and the phosphate leaving occur together. However, a key difference here is that the transition state no longer has a major proton transfer component to it, which can be seen by the fact that the base hydrogen distance is 0.99 \AA for Asp259 and 1.12 \AA in the phosphate base mechanism. This results in a modest barrier of 14.5 kJ mol^{-1} . From here, the reaction will collapse to a thermodynamically favourable product complex $200.7 \text{ kJ mol}^{-1}$ below the reactant complex.

This reaction pathway is much lower in energy than the phosphate mediated version, having a rate-limiting step 70.7 kJ mol^{-1} lower in energy. In fact the rate-limiting step of this reaction is not the thiol attack, rather it is a dihedral rotation. When the attack of the two thiols is compared, we see a reduction of 83.7 kJ mol^{-1} . The lowering of transition state energies indicates a likely role of Asp259 in the mechanism of Hcy self-cyclization.

4.6 Hse Self-Cyclization

It must be noted that even though the above mechanism proceeds with small energy barriers, it is not enough to infer Asp259's involvement in the cyclization reaction. Hcy is one of the most reactive amino acids and Hcy-AMP will readily self-cyclize in solution.^{23,24} Hse however is less reactive and provides a better measurement of Asp259's role in mediating cyclization. Therefore the cyclization mechanism for Hse in MetRS was also investigated. The results of the molecular dynamics simulations revealed large similarities in the positions of the active site residues in MetRS when complexed Hcy-AMP's and Hse-AMP indicated by RMSD of 0.41 Å. Molecular dynamics revealed average distance of the phosphate and carbonyl carbon with the hydroxyl of serine's side chain to be 6.76 Å and 3.79 Å, respectively. Furthermore, the average distance between Hse's hydroxyl and Asp259's oxygen is 5.21 Å. Although this is far, Asp259 undergoes a rotation during the simulation (figure 4.6a) resulting in the two-carboxylate oxygens switching positions this makes the 5.21 Å more representative as an average distance between both oxygen's. When only one conformer is examined the average distance becomes 4.28 Å. This rational of the rotation is further strengthened by examining the distance between the Asp259's carboxylate carbon and the Hse's hydroxyl oxygen (figure 4.6b), when this is examined we see a very consistent distance that averages to 5.65 Å.

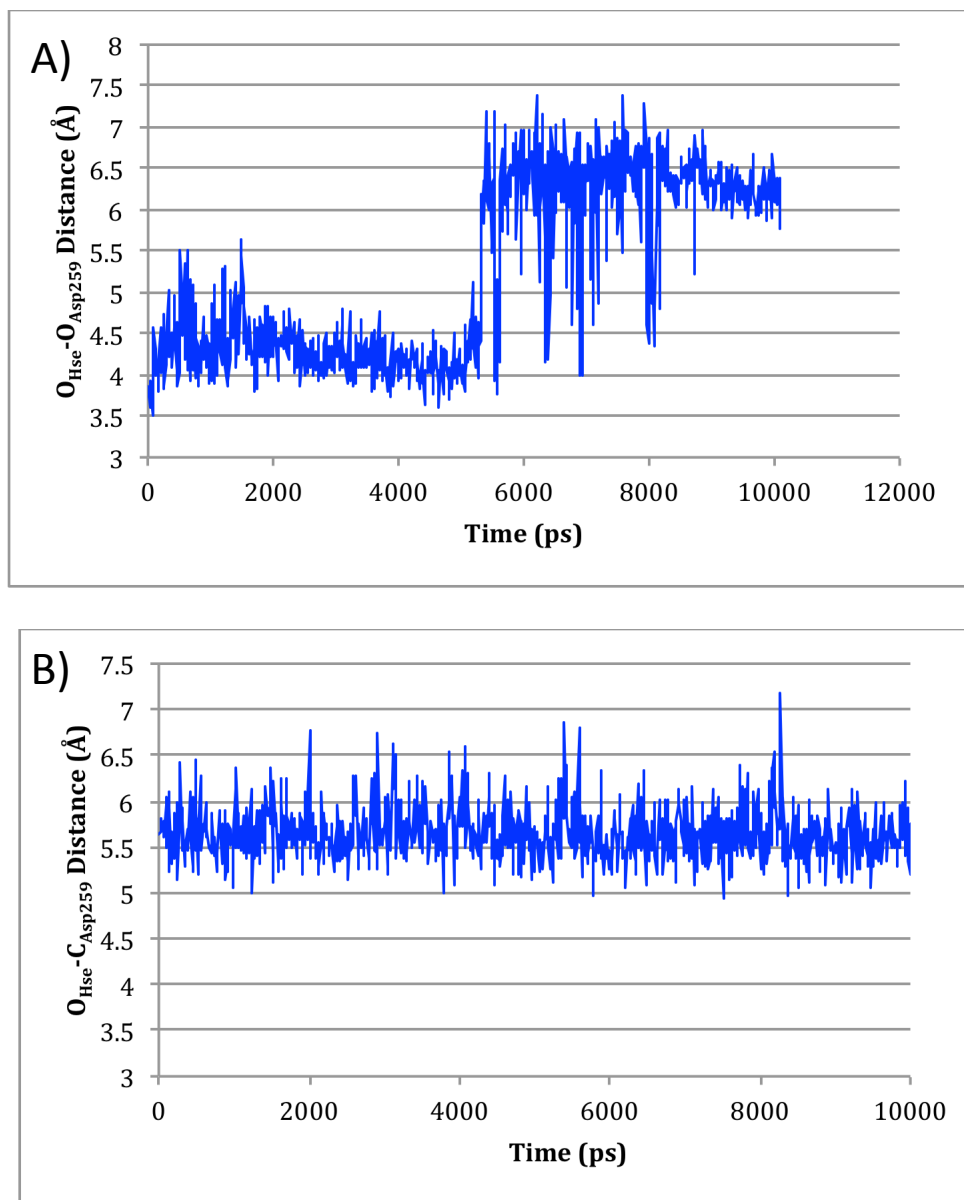
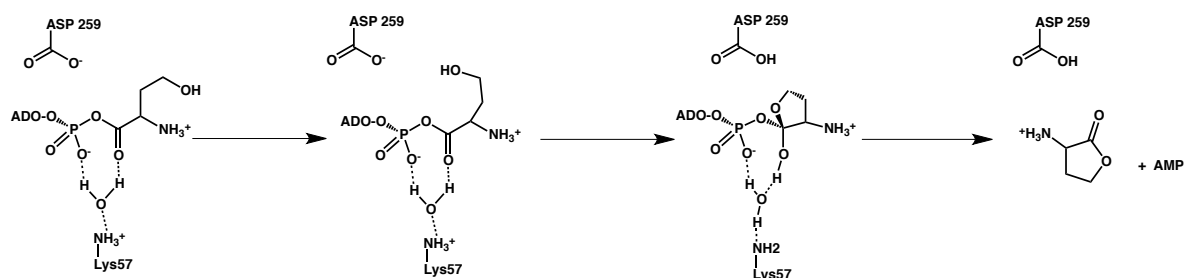


Figure 4.6. A) A graph showing the distance between the oxygen of Asp259 base and the Hse's oxygen during the molecular dynamics simulation. (At approximately 5 ns the carboxylate undergoes a rotation exchanging oxygen positions.) **B)** A graph showing the distance between the carbon of Asp259 base and the Hse's oxygen during the molecular dynamics simulation of MetRS.

The mechanism of Hse cyclization starts akin to Hcy. The reactive complex (figure 4.7) has the hydroxyl 3.59 Å away from the carbonyl carbon and 4.76 Å away from Asp259's oxygen. Hse then undergoes a dihedral rotation with a barrier of 19.1 kJ mol⁻¹ (scheme 4.2).

Scheme 4.2. An illustration a reaction pathway for the cyclization of Hse in MetRS utilizing Asp259 as a mechanistic base.



Once in this new conformation (figure 4.7), the hydroxyl was only 3.03 Å from the carbonyl carbon and engaged in a strong 1.83 Å hydrogen bond with Asp259. This complex is 39.1 kJ mol⁻¹ more favourable than the reactant complex.

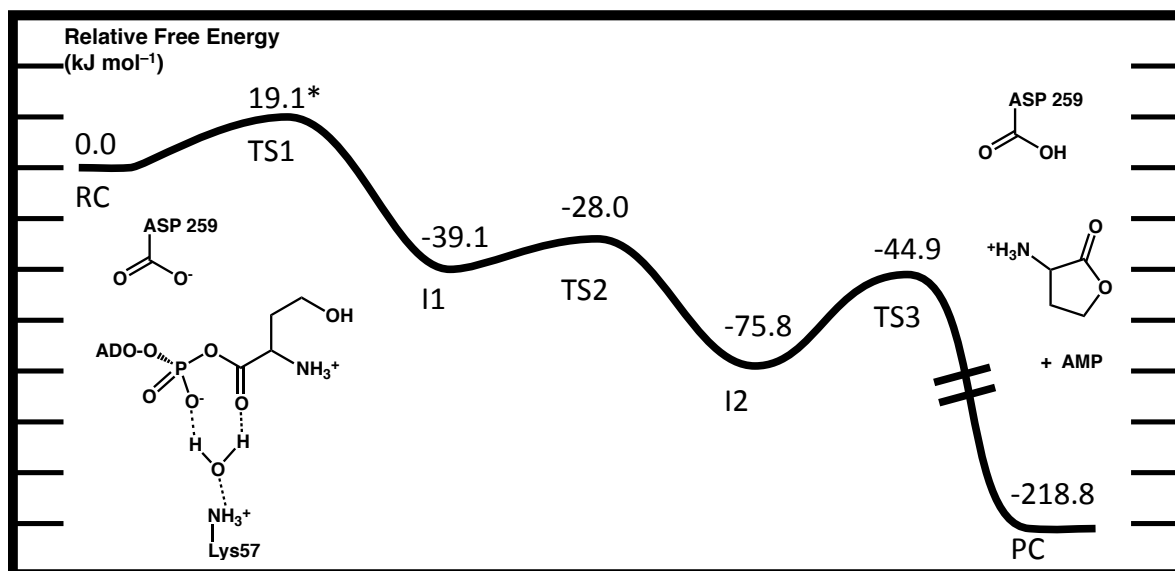


Figure 4.7. The potential energy surface for the cyclization reaction of Hse-AMP in MetRS with all relative energies reported in kJ mol⁻¹

The next transition state lies 11.1 kJ mol⁻¹ higher in energy than the current intermediate. It is characterized by the deprotonation of Hse by Asp259; the attack of the sp² carbon by the newly formed alkoxide and a proton is shuttled from Lys57 to the carbonyl oxygen via a water molecule forming an alcohol. This results in an sp³-hybridized carbon that is -75.8 kJ mol⁻¹ more stable than the reactant complex. The next step is rate-limiting. Here, the proton of the newly formed alcohol is shuttled via a bridging water molecule to the phosphate concomitantly as the phosphate is being cleaved from the newly formed lactone. The cleavage of the phosphate has a barrier of 30.9 kJ mol⁻¹ (slightly more than the rotation barrier for Hcy).

The transition state barriers for the cyclization of Hse-AMP are promising because this reaction is essentially the same one that occurs in the active site. An alcohol is deprotonated and attacks a carbonyl carbon releasing AMP. However rather than having typical barriers of over 100 kJ mol⁻¹, it accomplishes the task with less than an third of the energy.^{21,22} This significantly strengthens the argument that Asp259 plays a vital role in the editing process of Hcy and Hse by MetRS.

4.7 A General Motif of Hcy Editing

MetRS is not the only aaRS that has to edit against Hcy being incorporated into proteins. LeuRS, IleRS and ValRS will also readily bind Hcy.^{23,24} Thus, since these enzymes also bind Hcy, they are likely to have mechanisms by which they can facilitate its cyclization. To investigate this how this may be accomplished, we first examined the distance between the sulfur of Hcy's thiol and the non-bridging oxygen's of the phosphates (table 4.1, figure 4.8). When this is examined we see that for MetRS, ValRS and LeuRS the phosphate is far from the thiol and unlikely to be able to act as a base. IleRS on the other hand is in a favourable position to have its proton abstracted by AMP's phosphate. We then began the search for a possible base to abstract a proton. IleRS had no bases near the carbonyl carbon, which may indicate that in this enzyme the phosphate is the mechanistic base. In ValRS and LeuRS, Asp490 and Glu532 respectively were found to be in close proximity to the carbonyl carbon. These two residues are situated in a similar position as Asp259 in MetRS. In ValRS, the average distance between the thiol and Asp490's oxygen is 4.94 Å and the distance

between Asp490's oxygen and the sp^2 carbon is 5.26 Å. These distances are large. However, there is a water molecule that is hydrogen bonding to Asp490 (figure 4.8, table 4.1). This water could allow the thiol's proton to be shuttled over longer distances. In LeuRS, Glu532 is in a similar position to MetRS's Asp259. This carboxylate is close to the carbonyl carbon lying only 4.39Å away on average, close to the 4.25 Å of MetRS. The distance between the sulfur thiol of Hcy and Glu532 is 7.64 Å, which is similar to the distance to the phosphate base. However, like MetRS, if Hcy undergoes a dihedral rotation in LeuRS the sulfur is now 4.67 Å away from Glu532 while remaining 6.39 Å away from the phosphate. This now allows for Glu532 to sit in a very favourable position to facilitate the cyclization reaction.

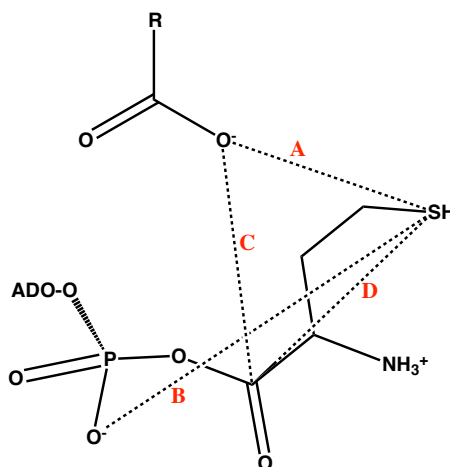


Figure 4.8. An illustration of Hcy and a carboxylate (either aspartic acid or glutamic acid) within the active site. Selected distances are highlighted A) $S_{Hcy}-O_{carboxylate}$ B) $S_{Hcy}-O_{phos}$ C) $O_{carboxylate}-C_{sp2}$ D) $S_{Hcy}-C_{sp2}$.

Table 4.1. An overview of selected average distances for Hcy bound in MetRS, LeuRS, ValRS and IleRS.

*Indicates that a bridging water is bound to the carboxylate

| Distance (Å) | MetRS | LeuRS | ValRS | IleRS |
|---|-------|-------|-------|-------|
| A. S_{Hcy}-O_{phos} | 6.99 | 7.78 | 7.01 | 4.55 |
| B. S_{Hcy}-O_{carboxylate} | 5.30 | 7.64 | 4.94 | - |
| C. S_{Hcy}-C_{sp2} | 3.95 | 5.19 | 4.14 | 4.57 |
| D. O_{carboxylate}-C_{sp2} | 4.25 | 4.39 | 5.26* | - |

4.8 Conclusions

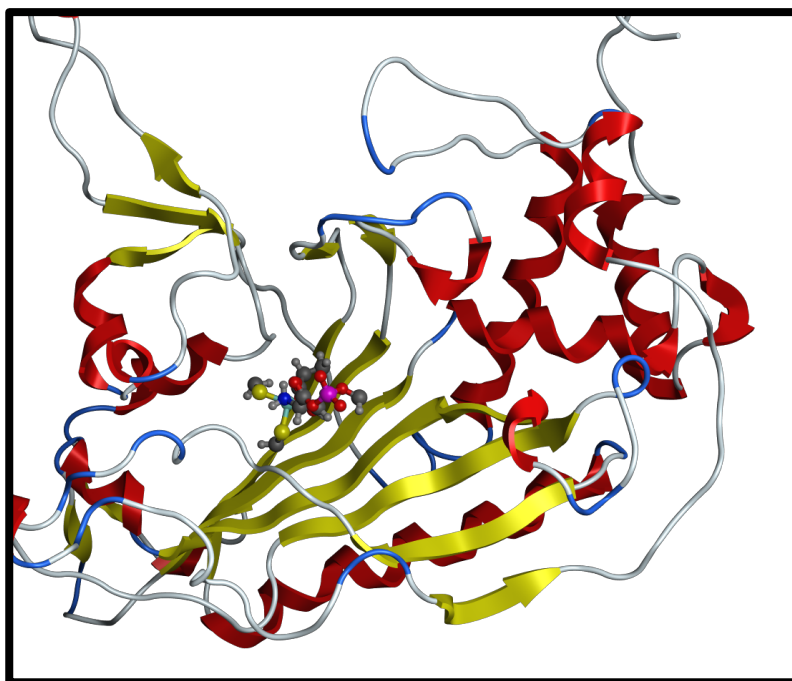
The cyclization reaction of Hcy in aaRS's is an important editing reaction that must be carried out successfully to maintain a healthy cell. MetRS is particularly efficient at accomplishing this, yet its mechanism is still unclear. When the typical phosphate base of aaRS's is used to facilitate the cyclization Hcy, a barrier of 98.2 kJ mol⁻¹ is calculated. Asp259 is an active site residue that remains close to the carbonyl carbon within the active site over molecular dynamics simulations and when used as a base is able to reduce the barrier of cyclization by over three fold. ValRS and LeuRS also have to edit against Hcy and have potential editing residues Asp490 and Glu532 respectively. This may indicate that different aaRS's may have evolved similar active site motifs in order to edit against highly reactive non-standard amino acids.

4.9 References

- (1) Hendrickson, T. L.; Schimmel, P. *Translation mechanisms* **2003**, 34.
- (2) Jakubowski, H. *Biochemistry* **1996**, 35, 8252.
- (3) Jakubowski, H. *Cell. Mol. Life Sci.* **2004**, 61, 470.
- (4) In *Molecular Operating Environment (MOE)*; Chemical Computing Group Inc: 010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7,, 2012.
- (5) M. J. Frisch, G. W. T., H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, and D. J. Fox; Gaussian, Inc.: Wallingford CT, 2010, p .
- (6) Bearpark, M. J.; Ogliaro, F.; Vreven, T.; Boggio-Pasqua, M.; Frisch, M. J.; Larkin, S. M.; Robb, M. A. In *Computation in Modern Science and Engineering Vol 2, Pts a and B*; Simos, T. E., Maroulis, G., Eds.; Amer Inst Physics: Melville, 2007; Vol. 2, p 583.
- (7) Dapprich, S.; Komaromi, I.; Byun, K. S.; Morokuma, K.; Frisch, M. J. *Theochem-J. Mol. Struct.* **1999**, 461, 1.
- (8) Humbel, S.; Sieber, S.; Morokuma, K. *J. Chem. Phys.* **1996**, 105, 1959.
- (9) Maseras, F.; Morokuma, K. *J. Comput. Chem.* **1995**, 16, 1170.

- (10) Svensson, M.; Humbel, S.; Froese, R. D. J.; Matsubara, T.; Sieber, S.; Morokuma, K. *J. Phys. Chem.* **1996**, *100*, 19357.
- (11) Vreven, T.; Byun, K. S.; Komaromi, I.; Dapprich, S.; Montgomery, J. A.; Morokuma, K.; Frisch, M. J. *J. Chem. Theory Comput.* **2006**, *2*, 815.
- (12) Vreven, T.; Morokuma, K. *J. Comput. Chem.* **2000**, *21*, 1419.
- (13) Vreven, T.; Morokuma, K.; Farkas, O.; Schlegel, H. B.; Frisch, M. J. *J. Comput. Chem.* **2003**, *24*, 760.
- (14) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.
- (15) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 1372.
- (16) Lee, C. T.; Yang, W. T.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.
- (17) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. *J. Comput. Chem.* **2005**, *26*, 1668.
- (18) Handy, N. C.; Cohen, A. J. *Mol. Phys.* **2001**, *99*, 403.
- (19) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623.
- (20) Cusack, S. *Curr. Opin. Struct. Biol.* **1997**, *7*, 881.
- (21) Huang, W. J.; Bushnell, E. A. C.; Francklyn, C. S.; Gault, J. W. *Journal of Physical Chemistry A* **2011**, *115*, 13050.
- (22) Liu, H. N.; Gault, J. W. *J. Phys. Chem. B* **2008**, *112*, 16874.
- (23) Jakubowski, H. *Acta Biochim. Pol.* **2011**, *58*, 149.
- (24) Jakubowski, H. *Wiley Interdiscip. Rev.-RNA* **2012**, *3*, 295.
- (25) Nakanishi, K.; Ogiso, Y.; Nakama, T.; Fukai, S.; Nureki, O. *Nat. Struct. Mol. Biol.* **2005**, *12*, 931.

Chapter 5

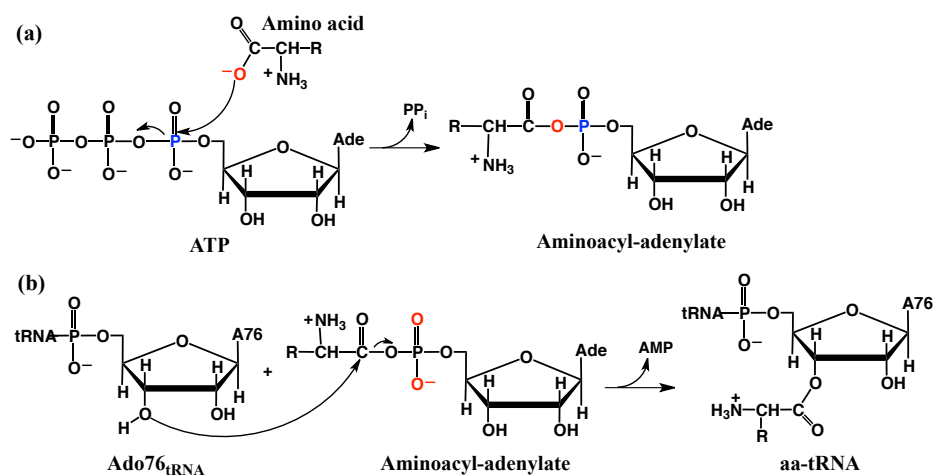


**A Computational Study into an Additional
Key Role of Zinc in SerRS: Inhibition of
Substrate Self-Cyclization**

5.1 Introduction

The importance carrying out protein synthesis correctly in biological systems cannot be overstated as proteins are fundamentally responsible for the survival of all known living organisms. This task requires that proteins have the correct number of amino acids as well as each being placed in the correct order.¹⁻³ If this not accomplished with high fidelity, several complications may arise including development of various diseases from cardiovascular disease to cancer^{4,5} Fortunately, nature has developed a method to ensure that the needed accuracy is achieved by the utilization aminoacyl tRNA synthetases (aaRSs). These enzymes catalyze the activation of amino acids as well as their sequential acylation (or transfer) to cognate tRNA (Scheme 5.1). These reactions are performed with remarkable fidelity indicated by errors only occurring once in every 10 000 reactions.¹ Furthermore, aaRS's are also required to edit non-standard amino acids, as well as work as molecular signals for processes such as apoptosis and even viral assembly.⁶⁻⁸

Scheme 5.1. A schematic illustration of a) the activation of amino acids and b) the transfer of amino acids to their cognate tRNA in protein synthesis.



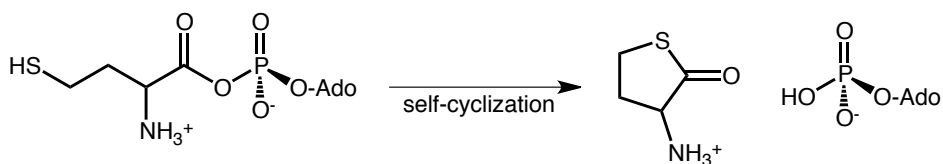
AaRSs can be used to probe the evolution of biological systems.⁹ Since aaRSs are thought to be amongst the first class of enzymes synthesized, we can evaluate when organisms diverged from each other by examining the differences between aaRSs from different organisms.¹⁰ Studies of aaRS's have suggested that serine, threonine and cysteine were most recently encoded into our genome.^{9,10} As a result, mammalian SerRS, ThrRS and CysRS can differ greatly from bacterial aaRS's and this creates the potential for novel antimicrobial pharmaceuticals to be developed that will selectively inhibit bacterial aaRSs but not mammalian.¹¹⁻¹³

One plausible reason for the delayed addition of Thr, Ser and Cys into our genome is that each of their respective aaRSs requires a Zn (II) ion, which is unique to these three aaRSs.¹⁴ The function of Zn (II) has been traditionally credited with selective and/or catalytic roles for aaRSs. In particular, the catalytic function of the Zn (II) has more recently been elucidated in ThrRS.¹⁵ Here, Zn (II) functions to increase the acidity of the α -amine, thus in result, exists in a neutral form. This allows the amine to act as the catalytic base in the acylation step of protein synthesis.^{15,16} Collectively, there are many candidates for bases that exist in nature: histidine, aspartic, glutamic acid and even the non bridging oxygen's of the AMP phosphate have been shown to act as bases in enzymatic catalysis.^{17,18} Zn (II)'s catalytic role could therefore be a byproduct of nature taking advantage of a metal ion that was already present in the active site since many other bases in nature could function akin to the deprotonated α -amine without needing Zn(II). Another proposed reason for why Zn (II) is in the active site of SerRS, CysRS and ThrRS is that of selectivity. Again, this role of Zn (II) is well studied and it has been shown that Zn (II) functions to provide selectivity in CysRS and SerRS as they do not require external editing site (a feature present in many aaRS's).^{14,19,20} However this selectivity can also be achieved in many other aaRSs such as TryRS and AspRS; they accomplish this without the use of an

external editing site nor a metal center.²¹⁻²⁵ In contrast, aaRSs that cannot select against non-cognitive amino acids with a high degree of accuracy have developed secondary editing mechanisms.²¹⁻²⁵ Finally, ThrRS uses both external editing site as well as a Zn (II) center within its active site.^{26,27} Thus, the correlation between selectivity, the presence of Zn(II) and an editing site is not clear.

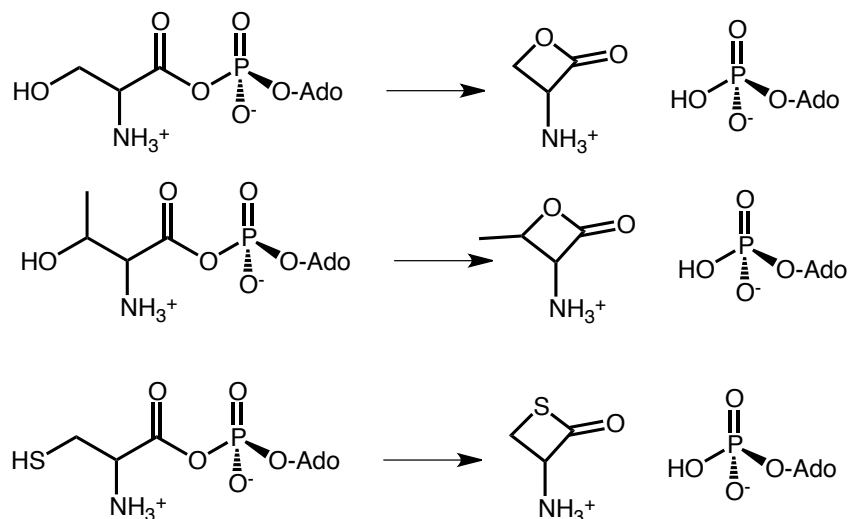
In an attempt enlighten these details of why Zn(II) is present we turn to MetRS because this aaRS is able to select against nonstandard homocysteine (Hcy) in addition to the naturally occurring amino acids.²⁸ This is important because Hcy can cause a number of physiological complications if incorporated into proteins such as neurodegenerative disorders and stroke.⁵ MetRS has developed a way to minimize this threat by capitalizing on a self-cyclization reaction (Scheme 5.2).⁵

Scheme 5.2. A schematic illustration highlighting the self-cyclization editing function in MetRS



When Hcy binds to MetRS, ATP will activate the Hcy. However, before Hcy-AMP can be acylated and transferred to tRNA, the free thiol group will attack the sp^2 carbon of the phosphoester functional group forming the cyclic product.²⁸ This cyclization has been shown to occur in solution for homoserine (Hse) as well as homocysteine. Therefore it is possible that Ser-AMP and Cys-AMP may be able to undergo the same reaction in the enzymes active site (Scheme 5.3).²⁸

Scheme 5.3. An illustration of possible side reactions that could take place in SerRS, ThrRS and CysRS.



A reason why this cyclization may not occur is due to the increased ring strain when going from the five membered Hcy and Hse to the strained four membered Cys, Ser and Thr products. However four membered rings are not only known to exist in nature (penicillin) and on the lab bench (beta lactones and lactams) but they can also be quite stable.^{29,30} This leads to the possibility that Zn (II) may be present to inhibit this cyclization by constraining on the binding of the hydroxyl or thiol groups.

Herein we hope to elucidate another key role of Zn (II) in the active sites of SerRS, CysRS and ThrRS in which it could act to inhibit a self-cyclization side reaction. To do this, we aim to use well established computational methods that have proven effective in evaluating enzymatic mechanisms.³¹ Using a synergistic approach of combining quantum mechanical (QM) clusters, molecular dynamic (MD) simulations and quantum mechanic and molecular mechanics (QM/MM) we hope to discern a possible origin of Zn (II) in aaRS's and to generalize a new role for metal ions in biological systems.

5.2 Computational Methods

5.2.1 Molecular Dynamics

The Molecular Operating Environment (MOE) program suite was used for model preparation and determination of the binding orientation of the substrate within the active site.³² The crystal structure of SerRS was taken from the protein data bank (PDB-ID: 3W3S).³³ This structure had serine-adenylate analog bound in the active site of SerRS, and was mutated *in silico* back to the native Ser-AMP. After these tasks were accomplished the missing hydrogen's were added using the MOE protonate 3D application. The geometry of each complex was then evaluated using the AMBER99 molecular mechanics force field until the root mean square deviation (RMSD) gradient of the energy fell below 0.01 kJ mol⁻¹ Å⁻¹. The MD simulations were performed under constant pressure and temperature. The equations of motion were then treated by NAMD, for which a time step for numerical integration was set to 0.5 fs. Initially the system was equilibrated for 0.1 ns at 300 K and then after its completion, a production run proceeded for 1 ns to obtain a conformational global minimum.³⁴ This approach has been previously used in multiple instances for similar systems.^{35,36}

5.2.2 QM/MM Calculations

All QM/MM calculations were performed using the ONIOM formalism within the Gaussian 09 program suite.³⁷⁻⁴⁵ Optimized structures were calculated at the ONIOM (B3LYP 6-31G(d,p):AMBER96) level of theory.⁴⁶⁻⁵¹ The substrate and catalytic residues within the active site were described using DFT and the remainder of the truncated model was described using the AMBER 96 molecular mechanics force field. Gibbs free energy corrections (at SATP, ΔE_{Gibbs}) were also calculated at this level of theory by computing the harmonic vibrational frequencies. In addition to the level of theory used to optimize

structures, a further stationary calculation was performed at all local minima and maxima along the reaction coordinate at ONIOM (B3LYP/6-311G(2df,p):AMBER96//B3LYP/6-31G(d,p):AMBER96)+ ΔE_{Gibbs} level of theory in the mechanical embedding (ME) formalism. In addition to using B3LYP single point calculations were done at ONIOM (M06/6-311G(2df,p):AMBER96//B3LYP/6-31G(d,p):AMBER96)+ ΔE_{Gibbs} and ONIOM (M06-2X/6-311G(2df,p):AMBER96//B3LYP/6-31G(d,p):AMBER96) + ΔE_{Gibbs} which have been shown to produce reliable energies.⁵²⁻⁵⁴

5.2.3 Model Preparation

The model used for the QM/MM calculations was taken from the calculations completed by the AMBER99 MD simulations from which a representative structure was selected and truncated. This truncation was completed by extending out from the substrate (Ser-AMP) so that there were three layers of amino acids that surrounded the substrate (i.e. the layer surrounding Ser-AMP, the layer surrounds that shell and finally one more layer surrounding all of the previous), which composed of over 2000 atoms. This many amino acids were included in the model because previous work has shown that long-range steric effects can have substantial influence on transition state energies.⁵⁵

The QM region of the model varied between the models used, in the native form the QM region consisted of Ser-AMP, Arg349 Arg366 which were thought to stabilize the phosphates negative charge, Zn (II) and its ligating residues Glu368, Cys478 and Cys319.

Two mutant versions were mutated *in silico*. In each model, the Zn (II) ion was deleted and Cys478 and Cys319 were put in the low layer, protonated and had their sulfurs fixed to ensure the active site maintained its original structure. These two structures differed in the protonation of the α -amine, the first being neutral and the second with the α -amine fully protonated as you would expect to find at physiological conditions. After the QM layer was

assigned the fixing was accomplished by expanding out 10 Å from all high layer atoms and allowing them to be mobile while anything further than that was fixed. Similar approaches have been used extensively and successfully in computational chemistry.⁵⁶

5.2.4 Quantum Mechanical Cluster

For the quantum mechanical clusters analogs of Ser-AMP, Thr-AMP and Cys-AMP, the AMP group was substituted for methylphosphate. These were used to gain insights into the intrinsic chemistry that occurs in self-cyclization at various dielectrics. This was carried out using IEFPCM-B3LYP/6-311G(2df,p)//IEFPCM-B3LYP/6-31G(d,p)+ ΔE_{Gibbs} , level of theory where the Gibbs corrections (ΔE_{Gibbs}) were obtained at the IEFPCM-B3LYP/6-31G(d,p) level of theory.^{37,46-51} The single point calculations were also preformed at IEFPCM-M06/6-311G(2df,p)//IEFPCM-B3LYP/6-31G(d,p)+ ΔE_{Gibbs} and IEFPCM-M06-2X/6-311G(2df,p)//IEFPCM-B3LYP/6-31G(d,p)+ ΔE_{Gibbs} to verify the B3LYP results.⁵²⁻⁵⁴

5.3. Results and Discussion

5.3.1 Quantum Mechanical Cluster Kinetics/Thermodynamics

The kinetics of self-cyclization of Cys-AMP, Thr-AMP and Ser-AMP are to our knowledge is currently unknown. This makes it difficult to predict if cyclization could occur in biological systems. Thus, in order to examine the possibility that this reaction may not occur at rates fast enough to cause issues in biological systems we studied the mechanisms of the self cyclization of the three amino acid adenylate moieties. In particular, we examined this reaction at dielectrics ranging from 1 to 78. This is important because we are studying charged species and they are known to be sensitive to their dielectric environment. Different mechanisms were observed when we examined the cyclization, for instance Ser and Thr cyclizes in a one step mechanism opposed to Cys, which occurs in a

two-step mechanism. However despite slightly varying mechanisms, a general trend still emerged that was an increase in barrier height as the dielectric increases (Figure 5.1). The only exception for this trend is Cys at the dielectric of 1 but this can be explained by the more unfavorable deprotonation of the thiol, creating a charged thiolate with no stabilization in gas phase.

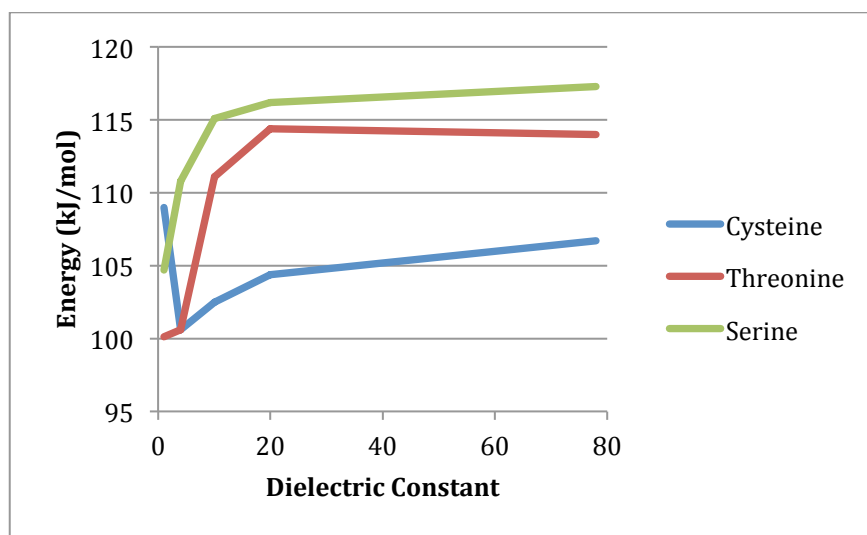


Figure 5.1. The rate-limiting step of the self cyclization reaction for cysteine, threonine and serine using IEF-PCM-B3LYP/6-311G(2df,p)//IEF-PCM-B3LYP/6-31G(d,p)+ ΔE_{Gibbs} .

For Thr and Ser the general trend can be easily explained. In the dielectric of 1 the anionic phosphate group is not well stabilized, thus during the proton transfer portion of the transition state, the greatest relative stabilization occurs when compared to other dielectric values. This explanation is further strengthened when the distance between the phosphate's oxygen and the hydroxyl hydrogen in the transition state is examined. This suggests the phosphate is most stabilized by the hydrogen at lower dielectric values (Figure 5.2) making cyclization easier to occur in lower dielectric environments such as proteins.

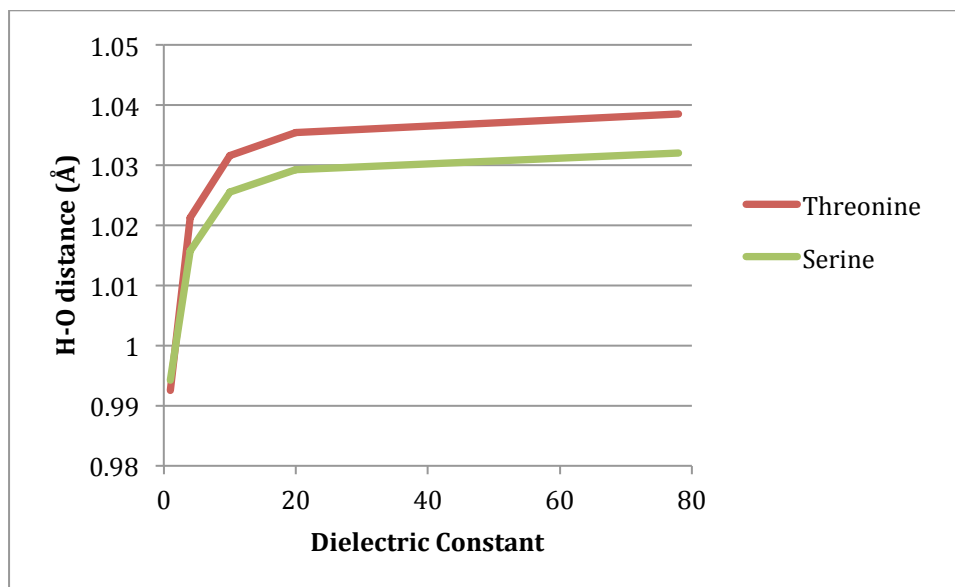


Figure 5.2. The bond distance with respect to dielectric of the attacking hydrogen from the hydroxyl group and the accepting phosphate's oxygen for the rate limiting transition state.

Interestingly Ser-AMP appears to have the highest barrier of cyclization regardless of functional (see Appendix) or dielectric value used. This reason for this is the opposite one would predict since secondary alcohols such as threonine generally have a higher pKa's than primary alcohols such as serine, consequently threonine's barrier should be higher. This suggests that a large component of the energetic barrier is not the deprotonation of the alcohol but the nucleophilic attack of the sp^2 which will favour stronger bases, in this case threonine. However it should be noted that even though serine does have the highest barrier it is still very close to the barriers calculated for the rate limiting acylation step of HisRS and ThrRS.^{15,17}

Based on the kinetic data presented above, we would predict that the self-cyclization reaction would occur more readily at low dielectric constants, but when the thermodynamics are examined we see the exact opposite trend (Figure 5.3).

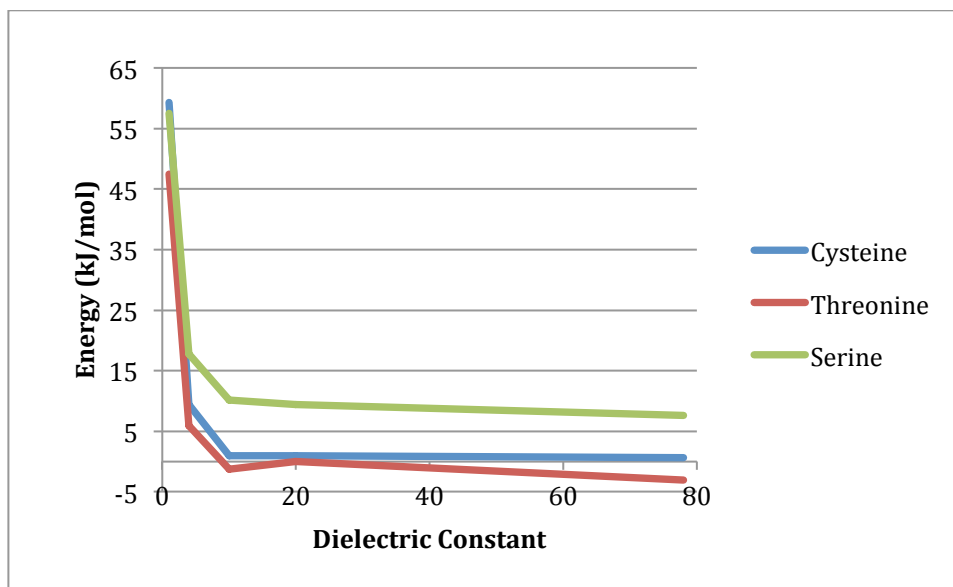


Figure 5.3. The overall thermodynamics of the self cyclization reaction for cysteine, threonine and serine using IEF-PCM-B3LYP/6-311G(2df,p)//IEF-PCM-B3LYP/6-31G(d,p)+ ΔE_{Gibbs}

For these results we see that the energies change when going from B3LYP to M06 and M06-2X. However the trend of the reaction is becoming more energetically favourable as the dielectric constant increases is conserved (see Appendix for M06 and M06-2X). This is in part due to the anionic phosphate-leaving group that is not sufficiently stabilized at low dielectric constants. Another result from the thermodynamic calculations is that Ser-AMP is consistently the least favourable self-cyclization reaction out of the systems studied. This means that for the three-aminoacyl adenylates studied, serine is least likely to undergo this reaction from both a kinetic and thermodynamic perspective. However, it is as if all three amino acids should be able to self-cyclize readily in both biological systems and aqueous ones. This means there should be some element present in CysRS, ThrRS and SerRS that is able to increase the barrier for this reaction so it does not occur.

5.3.2 QM/MM Model Choice

Ser-AMP in small models was the least likely to self-cyclize. Therefore, it was decided that SerRS should be chosen for our QM/MM model. This is because if the most difficult amino acid to self-cyclize can be shown to occur, then the cyclization of Cys and Thr may be able to be inferred from this data. Furthermore, SerRS has another advantage over CysRS and ThrRS in that for CysRS and ThrRS, both Cys and Thr ligate the Zn(II) metal center through their thiol and hydroxyl group respectively, whereas in SerRS, serine ligates the Zn(II) via only its α -amine.¹⁴ The ligation of Cys and Thr in CysRS and ThrRS allows for easy visualization of how the Zn(II) may inhibit self cyclization reaction since the Zn(II)-sulfur(oxygen) bond must be broken before the sp^2 carbon can be attacked. However in SerRS, there is no ligation of Zn(II) by the hydroxyl. This means that there should be more flexibility of the substrate increasing the likelihood of it being able to form the cyclic product. In light of this, serine seemed like the logical choice to investigate the potential inhibition role of Zn(II) in SerRS.

5.3.3 QM/MM without Zn(II)

The first model that will be discussed is the one in which the Zn(II) center was deleted and the ligating cysteine's were fixed and protonated. From there the system was allowed to minimize to the reactive complex (RC). In the RC, the Ser-AMP adenylate shifted so that the hydroxyl group of serine was no longer hydrogen bonding to Glu368. Instead, it was found to be involved in hydrogen bonding with the phosphate group of the AMP with a hydrogen bond distance of 2.58 Å. The α -amine then moved in to hydrogen bond with Glu368 forming a very strong hydrogen bond of 1.64 Å. This shift in hydrogen bonding allowed the hydroxyl group of serine to occupy a position to readily attack the sp^2 carbon center lying only 2.70 Å away from the carbon (Figure 5.4).

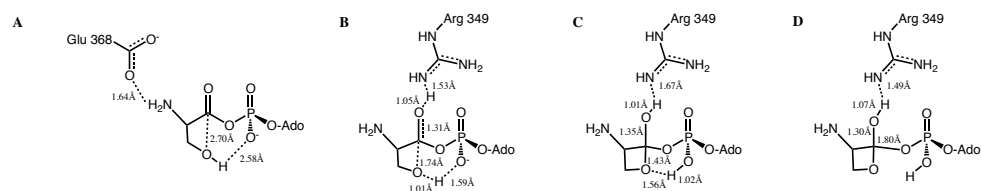


Figure 5.4. A) Optimized Reactive complex with selected bond distances (angstroms) highlighted. B) Optimized first transition state with selected bond distances (angstroms) highlighted. C) Optimized intermediate with selected bond distances (angstroms) highlighted. D) Optimized second transition state with selected bond distances (angstroms) highlighted.

From this point the phosphate began to abstract serine's hydroxyl hydrogen as the oxygen attacks the sp^2 carbon. Arg349 then donates a hydrogen to the forming oxyanion. This greatly stabilizes the transition state to a point, where when using B3LYP it lies 80.6 kJ mol^{-1} higher in energy than the RC, a huge improvement over the self-cyclization in solution (Figure 5.4). This result was expected, since the active site was designed to deprotonate a hydroxyl while allowing it to attack an sp^2 center, so it should be able to decrease the activation energy of such a reaction.

Upon completion, the complex relaxes to the intermediate that is 72.1 kJ mol^{-1} above the RC. In this state the bond between the attacking oxygen and carbon has been fully made along with Arg349's hydrogen and the oxyanion (Figure 5.4). In this intermediate the carbon phosphate oxygen bond has lengthened from 1.37 Å in the RC to 1.47 Å . This bond then continues to lengthen all the way to 1.80 Å as it enters the final transition state, which corresponds to the removal of AMP from serine (figure 5.4). This cleavage resulted in a low barrier only lying 1.6 kJ mol^{-1} higher in energy than the previous intermediate.

Following the cleavage of the carbon phosphate the product complex forms, which surprisingly is an exergonic process by 51.8 kJ mol^{-1} (figure 5.5). This is not an expected

result, as in the small models the vast majority of the reactions were endothermic, this extra release of energy can most likely be attributed to how well the phosphate is stabilized by the active site's arginine.

Single point calculations were also carried out using M06 and M06-2X (Appendix). Different absolute values for the barriers and overall thermodynamics are observed with these functionals. However, the barriers and thermodynamics when using M06 and B3LYP have all been reduced with respect to the small model calculations. Additionally, the new barriers calculated are lower than the calculated ones for the acylation reaction that should be taking place.^{15,17} This suggests that if the active site is without Zn(II) and its structure is maintained that self cyclization could pose a serious problem to SerRS.

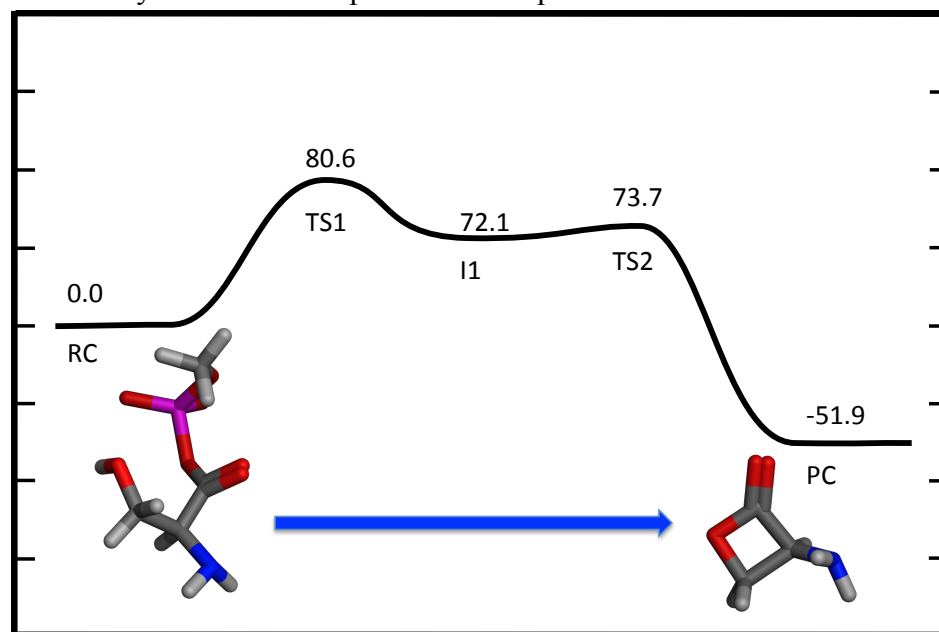


Figure 5.5. Calculated PES for the self cyclization reaction without Zn(II) and a neutral amine, B3LYP energies shown in kJ mol^{-1} .

5.3.4 QM/MM without Zn(II) and fully protonated α -amine

The above QM/MM results are promising but neglect the likely amines positive protonation state at physiological pH. Therefore, the next model used a protonated amine,

and observed major differences. In the reactive complex we observed the amine again move towards Glu368 however instead of hydrogen bonding like one would expect, the amine was deprotonated by Glu368, likely a consequence of the instability of charges in gas phase calculations. The amines donation of a proton to Glu368 causes subtle differences in the reactive complex shown in Figure 5.6. In this case we see a much stronger hydrogen bond between serine's hydroxyl and AMP's phosphate with a distance of 1.79 Å, but it comes at a tradeoff since the oxygen/ sp^2 carbon distance has now lengthened to 3.27 Å.

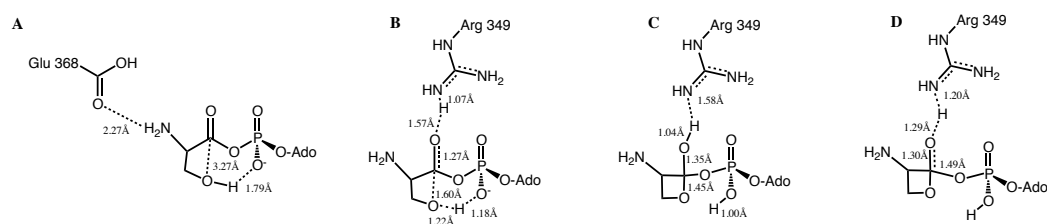


Figure 5.6. A) Optimized Reactive complex with selected bond distances (angstroms) highlighted. B) Optimized first transition state with selected bond distances (angstroms) highlighted. C) Optimized intermediate with selected bond distances (angstroms) highlighted. D) Optimized second transition state with selected bond distances (angstroms) highlighted.

This longer distance between the attacking oxygen and the sp^2 carbon becomes important as the reaction progresses to the first transition state, which now lies higher in energy at 91.1 kJ mol⁻¹ (Figure 5.6). The higher energy barrier could be because the attacking oxygen must now travel further, but it is likely a combination of that amongst other effects. Arg349 is not seen to stabilize the oxyanion as much as the previous model, which can be seen by the considerably longer bond length of 1.57 Å compared to 1.05 Å in the previous model. This decreased stabilization is evident in the increased double bond character of the amide oxygen carbon bond length of 1.27 Å compared to 1.31 Å in the previous model.

From this transition state, the complex relaxes 0.1 kJ mol^{-1} to its intermediate geometry. In this state we do not see many differences between this model and the previous, as all bond lengths are fairly similar as seen in 5. 4. The carbon-phosphate oxygen bond also lengthens from 1.36 \AA in the RC to 1.44 \AA similar to what was already seen. However as we move towards the second transition state, unlike the previous model (neutral amine) in which the bond lengthens to 1.80 \AA , this time the bond only lengthens to 1.49 \AA when it arrives at its second transition state (Figure 5.6). Once the second transition state is overcome the substrate collapses to a product complex that is more exergonic than the previous model, for all functionals examined (Figure 5.7).

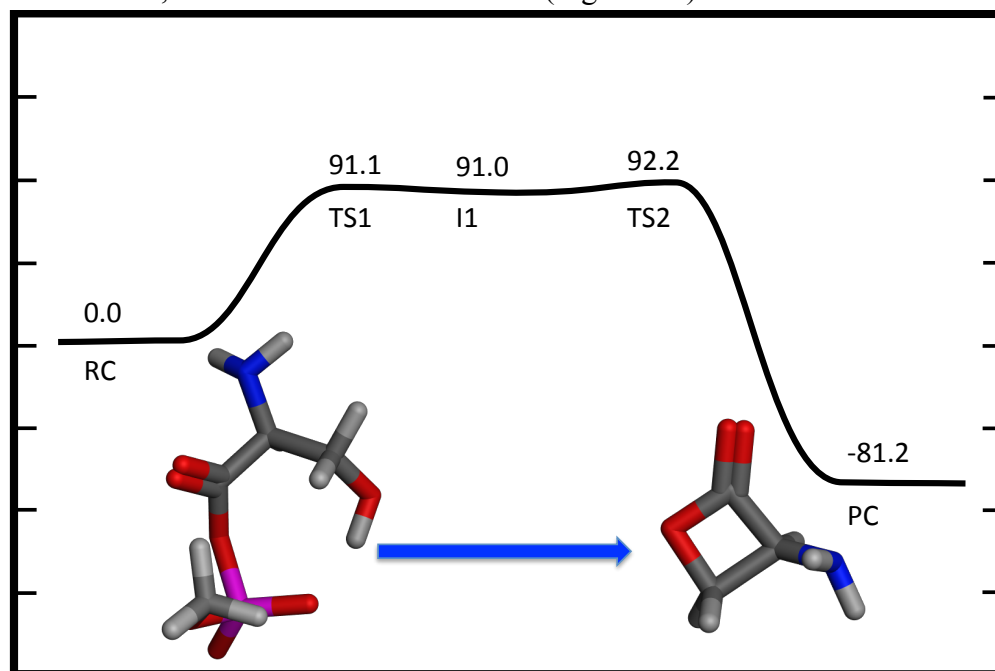


Figure 5.7. Calculated PES for the self cyclization reaction without Zn(II) and an protonated amine, B3LYP energies shown in kJ mol^{-1} .

5.3.5 QM/MM of Native SerRS

The final model that was investigated was the native enzyme with Ser-AMP present within the active site taken directly from the molecular dynamics simulation. We observed

some deviation from the crystal structure geometry during the course of the optimization. The most noteworthy difference was that the hydrogen bonding between the serine hydroxyl and Glu368 was no longer present. Instead Glu368 was found to hydrogen bond with Arg349 with a bond length of 1.50 Å, which is likely a consequence of the gas phase optimization. Instead of hydrogen bonding with Glu368 the hydroxyl was found to hydrogen bond with the thiolate of Cys478 ligating the Zn(II) ion. This results in a reactive complex where the hydroxyl oxygen is 3.66 Å away from its carbon and 5.42 Å away from the phosphate. From the reactive complex the hydroxyl once again moves towards the sp^2 carbon of Ser-AMP until it is 1.66 Å away, at which point it is at its first transition state (Figure 5.8), which lies 182.3 kJ mol⁻¹ above the reactive complex.

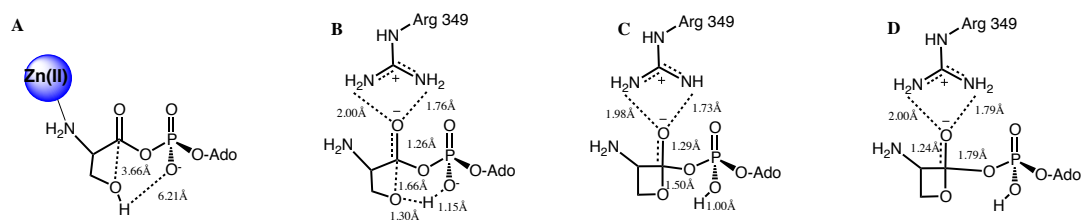


Figure 5.8. A) Optimized Reactive complex with selected bond distances (angstroms) highlighted. B) Optimized first transition state with selected bond distances (angstroms) highlighted. C) Optimized intermediate with selected bond distances (angstroms) highlighted. D) Optimized second transition state with selected bond distances (angstroms) highlighted.

This large difference in energy could be a consequence of several nuances within the active site. The first reason is that the hydroxyl sp^2 distance in the RC is 3.66 Å, which is longer than either of the previous models. However it is unlikely that this is the main cause for the doubling of the reaction barrier because in the previous models the hydroxyl- sp^2 -carbon distance is 2.70 Å and 3.27 Å and even with this large difference in distance the observed change in barrier height is not large. Another potential reason for this high barrier

is that Arg349 is not able to stabilize the oxyanion as well as in the previous two models. However this is unlikely the main cause for the high barrier, since in the small model calculations we saw the barrier for Ser-AMP to cyclize in gas phase ranged from 101.1 kJ mol⁻¹ to 104.7 kJ mol⁻¹ depending on the functional used, and being in gas phase, the oxyanion did not receive stabilization from the environment. A more likely reason for this high barrier is a steric one that is enforced by the Zn(II) α -amine bond not allowing the system to achieve its desired geometry. One of likely geometry constraints Zn(II) imposes is the orientation in which the phosphate binds. For the Zn(II) model the non bridging oxygen of the phosphate and hydroxyl of Ser are on different sides of the sp² carbon plane. However with the other models they lie on the same plane. The latter orientation makes it very difficult for serine to access the phosphate base, which could in turn cause the high barrier observed.

Once this large barrier has been overcome the reaction arrives at its first and only intermediate, which lies 179.9 kJ mol⁻¹ above the reactive complex. The reason for this high-energy value is likely driven by the same steric hindrances that seem to be at fault for the unfeasible transition state before it. Upon relaxing to this intermediate the phosphate's oxygen-carbon bond then lengthens to 1.79 Å in its second transition state compared to 1.51 Å in the intermediate and 1.31 Å in the reactive complex. Once this critical bond length is reached the system relaxes to the product complex, now 64.5 kJ mol⁻¹ above the reactive complex. Surprisingly this process is endothermic unlike the other two systems, which are exergonic.

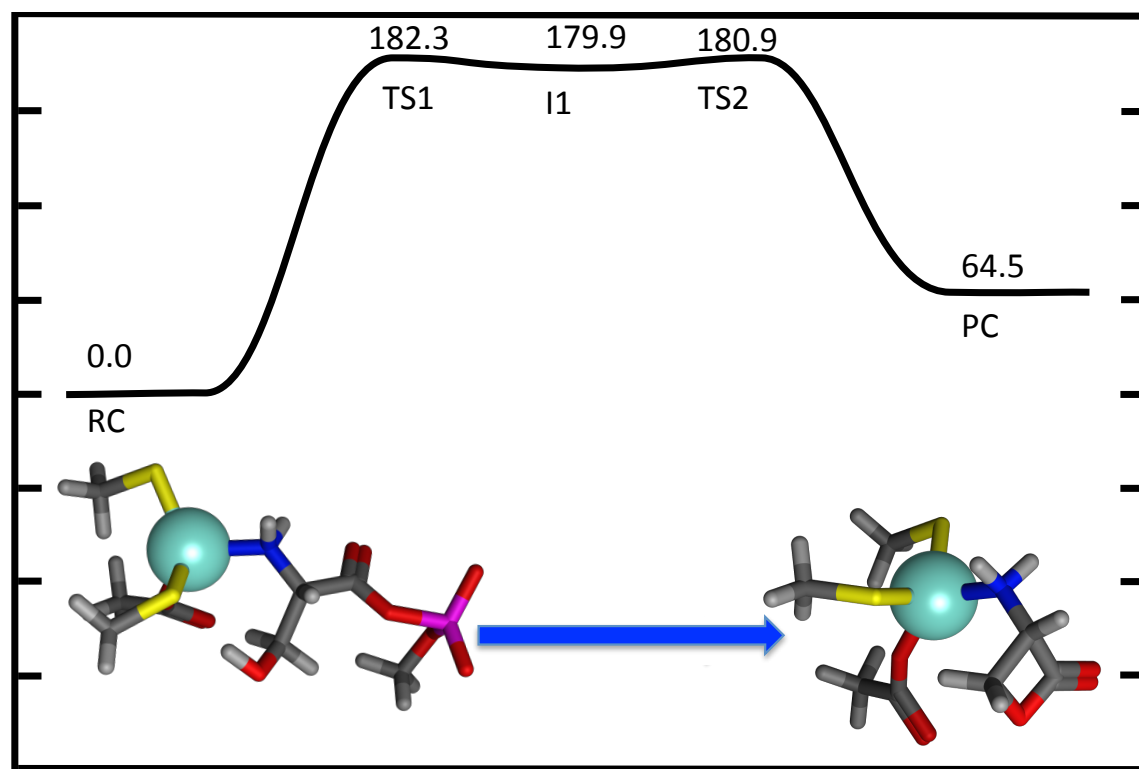
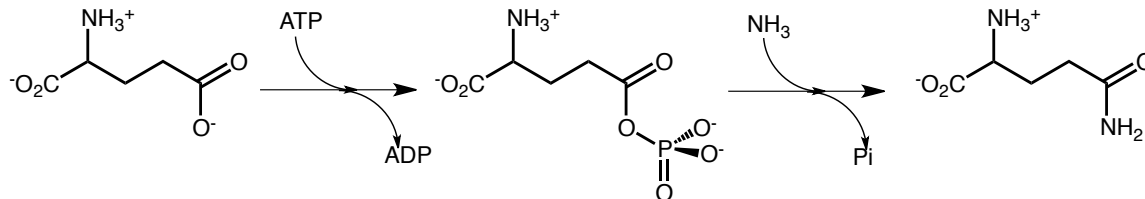


Figure 5.9. Calculated PES for the self cyclization reaction in the native enzyme, B3LYP energies shown in kJ mol⁻¹

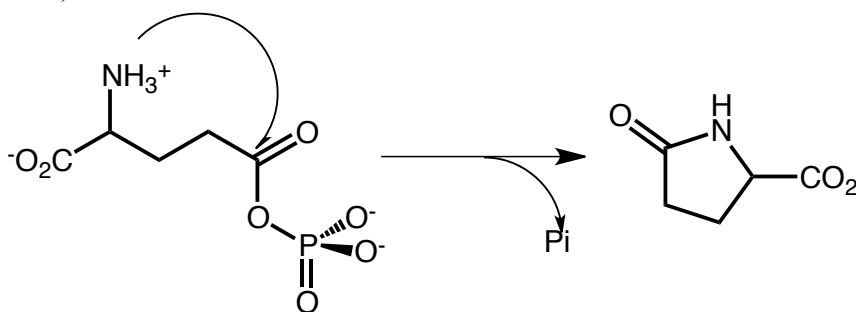
5.4 Applications of Current Work

Our results suggest that a role Zn (II) in aaRSs is to inhibit a self-cyclization side reaction inactivating the substrate. However metal ions in enzymes other than aaRSs should be able, in principal, to also inhibit cyclization. Also other metal ions could be able to have the same function. A current search of the literature reveals a potential answer to both of the points in the form of glutamine synthetase. This enzyme catalyzes the conversion of glutamate to glutamine (Scheme 5.4).⁵⁷



Scheme 5.4. The mechanism by which glutamine synthetase converts glutamate to glutamine.

However it is curious why the alpha amine cannot attack the sp^2 carbon and form a cyclic product (Scheme 5.5).



Scheme 5.5. A possible self-cyclization reaction that may occur in the active site.

One possible reason for this is elucidated by a crystal structure of Glutamine Synthetase (PDB IB:2LGS).⁵⁸ In this structure we see that glutamate is bound to a Mn(II) by its α -amine. This ligation could be in place to inhibit the self-cyclization. Thus, we again turned to a QM cluster approach and modeled the self-cyclization and varying dielectrics and with different functionals (Figures 5.10 and 5.11).

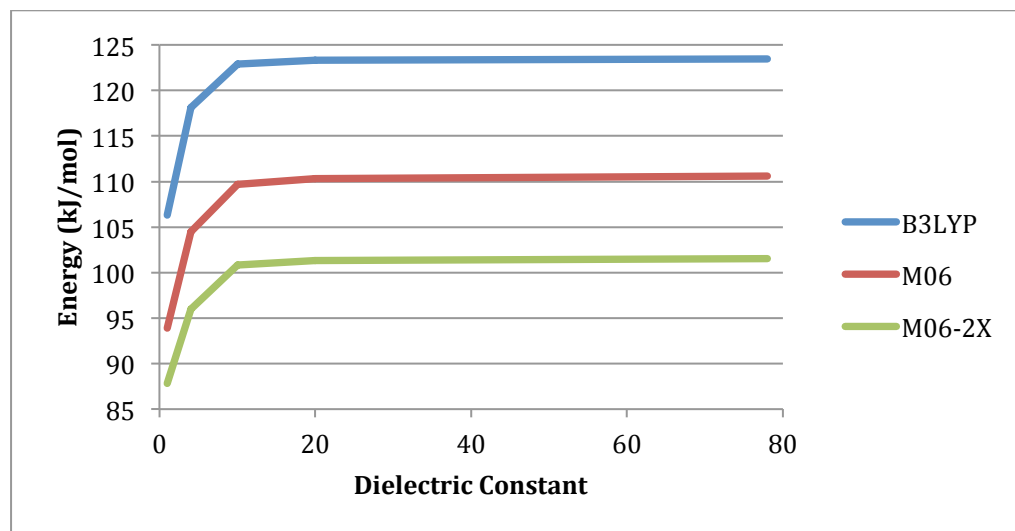


Figure 5.10. The rate-limiting step of the self cyclization reaction for glutamate phosphate using IEF-PCM-B3LYP/6-311G(2df,p)//IEF-PCM-B3LYP/6-31G(d,p)+ ΔE_{Gibbs} , IEF-PCM-M06/6-311G(2df,p)//IEF-PCM-B3LYP/6-31G(d,p)+ ΔE_{Gibbs} , IEF-PCM-M06-2X/6-311G(2df,p)//IEF-PCM-B3LYP/6-31G(d,p)+ ΔE_{Gibbs} .

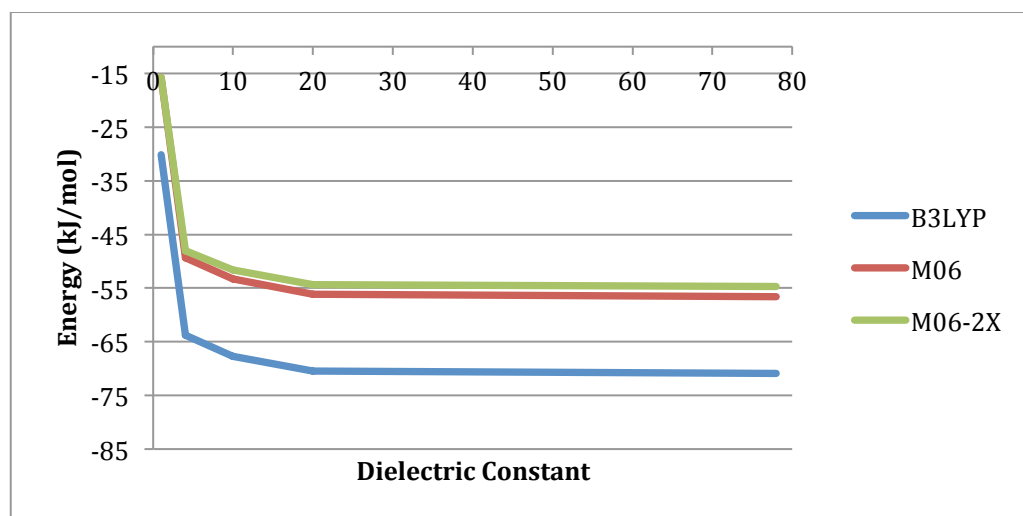


Figure 5.11. The overall thermodynamics of the self cyclization reaction for glutamate phosphate using IEF-PCM-B3LYP/6-311G(2df,p)//IEF-PCM-B3LYP/6-31G(d,p)+ ΔE_{Gibbs} , IEF-PCM-M06/6-311G(2df,p)// IEF-PCM-B3LYP/6-31G(d,p)+ ΔE_{Gibbs} , IEF-PCM-M06-2X/6-311G(2df,p)//IEF-PCM-B3LYP/6-31G(d,p)+ ΔE_{Gibbs} .

These results are analogous to the ones seen before with the barrier height predicted to be between 87.8 kJ mol⁻¹ and 123.5 kJ mol⁻¹ depending on the functional and dielectric environment used. However one difference between the self-cyclization of phosphoglutamate and Ser-AMP, Thr-AMP and Cys-AMP is that for all dielectric environments and functional used, we see that this process is exergonic. This release of energy is likely a byproduct of the reduced ring strain when going from a four membered ring to a five membered ring and potentially makes this possible side reaction even more debilitating than the ones investigated earlier.

5.5 Conclusions

In this work quantum mechanical clusters were used in conjunction with larger quantum mechanics/ molecular mechanics to investigate a new role of Zn(II) in tRNA synthetases. In particular, how Zn (II) may aid in the inhibition of a self-cyclization side reaction in ThrRS, SerRS and CysRS, and how this function may be generalized to other biochemical systems.

The quantum mechanical clusters gave insights into the energetics and kinetics for the self-cyclization reaction involving Cys-AMP, Ser-AMP and Thr-AMP. In all cases examined, the barriers were feasible. In addition, Thr-AMP can be compared to the energetics of the actual second step, for which the difference is less than 2 kJ mol⁻¹ (at protein dielectric values). This implies that the reaction is not only feasible but also likely to compete with the desired reaction.

QM/MM calculations provided insights on how readily this self cyclization side reaction could occur inside SerRS without Zn(II) present. Calculated barriers were lower than the actual acylation reaction that should be taking place within the active site. However, when Zn(II) is present within the active site, it enforces steric restraints and a

doubling of the barrier height to $182.3 \text{ kJ mol}^{-1}$ is observed, making the self cyclization reaction difficult. This suggests that Zn(II) is a necessary component needed to inhibit this side reaction.

Furthermore the possibility of other metals in other enzymes carrying out similar functions was examined by looking at a potential self-cyclization side reaction of glutamine synthetase using a quantum mechanical cluster. Using this approach we have concluded that unless otherwise restrained, a self-cyclization reaction could occur with similar kinetics and more favourable thermodynamics. This finding lends itself to suggest that the Mn(II) ion in the active site of glutamine synthetase may also carry out a similar function to that examined in SerRS.

5.6 References

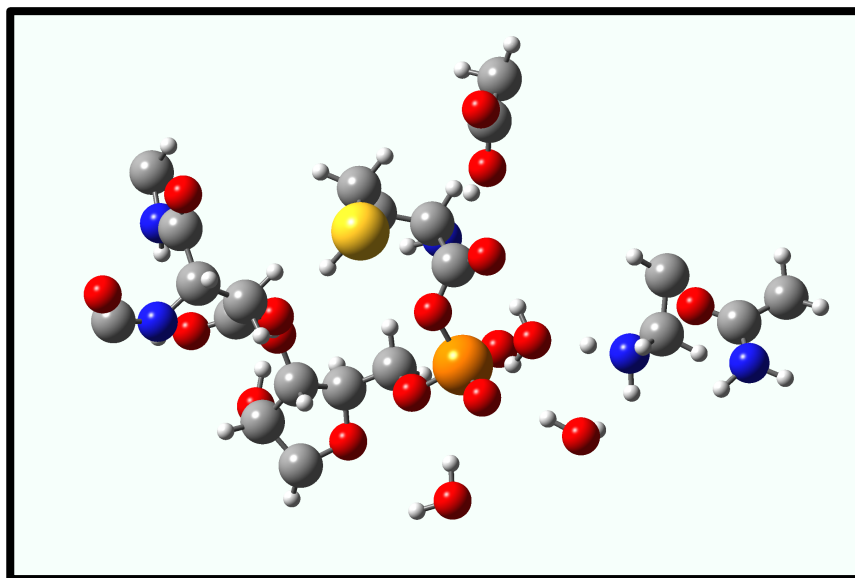
- (1) Cusack, S. *Curr. Opin. Struct. Biol.* **1997**, 7, 881.
- (2) Arnez, J. G.; Moras, D. *Trends in Biochemical Sciences* **1997**, 22, 211.
- (3) Safro, M. G.; Moor, N. A. *Molecular Biology* **2009**, 43, 211.
- (4) Sauls, D. L.; Warren, M.; Hoffman, M. *Thromb. Res.* **2011**, 127, 576.
- (5) Jakubowski, H. *Cell. Mol. Life Sci.* **2004**, 61, 470.
- (6) Hausmann, C. D.; Ibba, M. *Fems Microbiol. Rev.* **2008**, 32, 705.
- (7) Wang, J.; Fang, P. F.; Schimmel, P.; Guo, M. *J. Phys. Chem. B* **2012**, 116, 6798.
- (8) Dewan, V.; Liu, T.; Chen, K. M.; Qian, Z. Q.; Xiao, Y.; Kleiman, L.; Mahasenan, K. V.; Li, C. L.; Matsuo, H.; Pei, D. H.; Musier-Forsyth, K. *Acs Chemical Biology* **2012**, 7, 761.
- (9) Fournier, G. P.; Andam, C. P.; Alm, E. J.; Gogarten, J. P. *Orig. Life Evol. Biosph.* **2011**, 41, 621.
- (10) Fournier, G. P.; Andam, C. P.; Alm, E. J.; Gogarten, J. P. *Orig. Life Evol. Biosph.* **2012**, 42, 377.

- (11) Agarwal, V.; Nair, S. K. *Medchemcomm* **2012**, 3, 887.
- (12) Hurdle, J. G.; O'Neill, A. J.; Chopra, I. *Antimicrob. Agents Chemother.* **2005**, 49, 4821.
- (13) Bernier, S.; Dubois, D. Y.; Therrien, M.; Lapointe, J.; Chenevert, R. *Bioorg. Med. Chem. Lett.* **2000**, 10, 2441.
- (14) Bilokapic, S.; Maier, T.; Ahel, D.; Gruic-Sovulj, I.; Soll, D.; Weygand-Durasevic, I.; Ban, N. *Embo J.* **2006**, 25, 2498.
- (15) Huang, W. J.; Bushnell, E. A. C.; Francklyn, C. S.; Gauld, J. W. *Journal of Physical Chemistry A* **2011**, 115, 13050.
- (16) Bushnell, E. A. C.; Huang, W. J.; Llano, J.; Gauld, J. W. *J. Phys. Chem. B* **2012**, 116, 5205.
- (17) Liu, H. N.; Gauld, J. W. *J. Phys. Chem. B* **2008**, 112, 16874.
- (18) Dokainish, H. M.; Gauld, J. W. *Biochemistry* **2013**, 52, 1814.
- (19) Zhang, C. M.; Christian, T.; Newberry, K. J.; Perona, J. J.; Hou, Y. M. *J. Mol. Biol.* **2003**, 327, 911.
- (20) Newberry, K. J.; Hou, Y. M.; Perona, J. J. *Embo J.* **2002**, 21, 2778.
- (21) Kumar, S.; Das, M.; Hadad, C. M.; Musier-Forsyth, K. *J. Phys. Chem. B* **2012**, 116, 6991.
- (22) Pasman, Z.; Robey-Bond, S.; Mirando, A. C.; Smith, G. J.; Lague, A.; Francklyn, C. S. *Biochemistry* **2011**, 50, 1474.
- (23) Hussain, T.; Kamarthapu, V.; Kruparani, S. P.; Deshmukh, M. V.; Sankaranarayanan, R. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, 107, 22117.
- (24) Jakubowski, H. *Biochemistry* **1999**, 38, 8088.
- (25) Jakubowski, H.; Goldman, E. *Microbiol. Rev.* **1992**, 56, 412.

- (26) Sankaranarayanan, R.; Dock-Bregeon, A. C.; Rees, B.; Bovee, M.; Caillet, J.; Romby, P.; Francklyn, C. S.; Moras, D. *Nature Structural Biology* **2000**, 7, 461.
- (27) Minajigi, A.; Francklyn, C. S. *J. Biol. Chem.* **2010**, 285, 23810.
- (28) Jakubowski, H. *Acta Biochim. Pol.* **2011**, 58, 149.
- (29) Garud, D. R.; Makimura, M.; Koketsu, M. *New J. Chem.* **2011**, 35, 581.
- (30) Han, S.; Caspers, N.; Zaniwski, R. P.; Lacey, B. M.; Tomaras, A. P.; Feng, X. D.; Geoghegan, K. F.; Shanmugasundaram, V. *J. Am. Chem. Soc.* **2011**, 133, 20536.
- (31) Llano, J.; Gauld, J. W. In *Quantum Biochemistry*; Wiley-VCH Verlag GmbH & Co. KGaA: 2010, p 643.
- (32) *Molecular Operating Environment (MOE)*; Chemical Computing Group Inc: 010 .
- (33) Itoh, Y.; Sekine, S. I.; Suetsugu, S.; Yokoyama, S. *Nucleic Acids Res* **2013**.
- (34) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. *J. Comput. Chem.* **2005**, 26, 1781.
- (35) Bushnell, E. A. C.; Erdtman, E.; Llano, J.; Eriksson, L. A.; Gauld, J. W. *J. Comput. Chem.* **2011**, 32, 822.
- (36) Ion, B. F.; Bushnell, E. A. C.; De Luna, P.; Gauld, J. W. *Int. J. Mol. Sci.* **2012**, 13, 12994.
- (37) Frisch, M. J.; et al. Gaussian 09, revision D.02; Gaussian, Inc.: Wallingford, CT, 2009
- (38) Bearpark, M. J.; Ogliaro, F.; Vreven, T.; Boggio-Pasqua, M.; Frisch, M. J.; Larkin, S. M.; Robb, M. A. In *Computation in Modern Science and Engineering Vol 2, Pts a and B*; Simos, T. E., Maroulis, G., Eds.; Amer Inst Physics: Melville, 2007; Vol. 2, p 583.
- (39) Dapprich, S.; Komaromi, I.; Byun, K. S.; Morokuma, K.; Frisch, M. J. *Theochem-J. Mol. Struct.* **1999**, 461, 1.
- (40) Humbel, S.; Sieber, S.; Morokuma, K. *J. Chem. Phys.* **1996**, 105, 1959.

- (41) Maseras, F.; Morokuma, K. *J. Comput. Chem.* **1995**, *16*, 1170.
- (42) Svensson, M.; Humbel, S.; Froese, R. D. J.; Matsubara, T.; Sieber, S.; Morokuma, K. *J. Phys. Chem.* **1996**, *100*, 19357.
- (43) Vreven, T.; Byun, K. S.; Komaromi, I.; Dapprich, S.; Montgomery, J. A.; Morokuma, K.; Frisch, M. J. *J. Chem. Theory Comput.* **2006**, *2*, 815.
- (44) Vreven, T.; Morokuma, K. *J. Comput. Chem.* **2000**, *21*, 1419.
- (45) Vreven, T.; Morokuma, K.; Farkas, O.; Schlegel, H. B.; Frisch, M. J. *J. Comput. Chem.* **2003**, *24*, 760.
- (46) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648.
- (47) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 1372.
- (48) Lee, C. T.; Yang, W. T.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785.
- (49) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. *J. Comput. Chem.* **2005**, *26*, 1668.
- (50) Handy, N. C.; Cohen, A. J. *Mol. Phys.* **2001**, *99*, 403.
- (51) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623.
- (52) Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2008**, *120*, 215.
- (53) Valero, R.; Gomes, J. R. B.; Truhlar, D. G.; Illas, F. *J. Chem. Phys.* **2008**, *129*.
- (54) Zhao, Y.; Truhlar, D. G. *Accounts Chem. Res.* **2008**, *41*, 157.
- (55) Hu, L. H.; Soderhjelm, P.; Ryde, U. *J. Chem. Theory Comput.* **2011**, *7*, 761.
- (56) Sousa, S. F.; Fernandes, P. A.; Ramos, M. J. *Phys. Chem. Chem. Phys.* **2012**, *14*, 12431.
- (57) Liaw, S. H.; Kuo, I. C.; Eisenberg, D. *Protein Sci.* **1995**, *4*, 2358.
- (58) Liaw, S. H.; Pan, C.; Eisenberg, D. *Proc. Natl. Acad. Sci. U. S. A.* **1993**, *90*, 4996.

Chapter 6



Conclusions

6.1 Conclusions

In this thesis a wide array of computational chemistry techniques have been employed to gain greater insights into the editing mechanisms of aaRS, and their accuracy. These investigations proved valuable in describing and characterizing these enzymes and elucidating some of the fundamental chemistry by which they operate.

In Chapters 3 and 4 the joint computational methodologies of molecular dynamics (MD) and quantum mechanics/molecular mechanics (QM/MM) were used to investigate the editing of Homocysteine (Hcy) in aaRSs. In particular, the ability of phosphate or carboxylic acid moieties to aid in the cyclization of Hcy was considered. We examined the binding orientations of Hcy-AMP, Hse-AMP and Met-AMP within MetRS and observed no major difference in either the binding orientation or active site conformation. It was also observed that for the enzymes MetRS, LeuRS and ValRS the phosphate of the AMP moiety of the aa-AMP substrate was not in a position that would be conducive for it to act as a base and facilitate the cyclization reaction. These three enzymes did have a conserved carboxylic acid moiety that was well-positioned to abstract a proton during cyclization. The role of the carboxylic acid of the Asp259 residue in MetRS was investigated in great detail via a QM/MM approach. Importantly, the Asp259 was able to reduce the barrier of cyclization by a factor greater than three, relative to an alternate mechanism in which the substrate's AMP's phosphate acts as the required mechanistic base. This outstanding rate enhancement may be enough to allow Hcy self-cyclization to occur at a rate that does not allow the acylation reaction to compete, effectively meaning no Hcy may be transferred to tRNA.

In Chapter 5 the role of Zn(II) in CysRS, ThrRS and SerRS was investigated, beginning with the use of quantum mechanical (QM)-chemical cluster approaches. More specifically, the potential for the Cys-AMP, Thr-AMP and Ser-AMP to self-cyclize

within their respective cognate aaRS's active site, analogous to Hcy-AMP, was examined. The results obtained suggest that all three activated amino acids have the potential to self-cyclize with barriers similar to those for aminoacyl transfer previously calculated for HisRS and ThrRS.^{1,2} This could be a serious problem if the reactants were able to deactivate themselves. QM/MM models were used to further investigate the possibility of self-cyclization occurring within the active site of SerRS both with the Zn(II) ion present and without it. When Zn(II) is not present in the active site, the calculated barriers were notably lower than those for aminoacyl transfer as catalyzed by HisRS and ThrRS.^{1,2} However, when the Zn(II) is present the barrier for self-cyclization is calculated to be significantly higher, at over 180 kJ mol⁻¹. Hence, when Zn(II) is involved in binding the substrate this alternate and undesirable reaction is unlikely to be feasible. This result suggests a previously unrecognised role for Zn(II) in the active sites of CysRS, SerRS and ThrRS and possibly other metalloenzymes.

Although this thesis is by no means an all-inclusive work on the accuracy and mechanisms of editing within aaRSs, it does provide valuable insights. Importantly it adds a great deal of knowledge to a field of aaRSs. This field is becoming a target for drug development and the computational methods employed in this thesis are becoming the methods of choice for companies or research groups involved in this area.³⁻⁶ This is because transition state-analogs are often the best inhibitors and computational chemistry offers an effective way to ascertain such structures, making it a truly valuable pursuit.

Future extensions of this work could be to examine mechanisms of how the self-cyclization reaction for Hcy and Hse occurs within the aminoacyl transfer active sites of IleRS, LeuRS and ValRS. The effects of a mutation of the proposed carboxylic acid to

an amide, or a methyl as in the case of alanine, could also be examined via MD simulations as well as QM/MM models. This would enable better prediction of the results experimentalists that are likely to be observed. These studies could be repeated for aaRS enzymes found in different organisms as well to examine the universality of this editing mechanism. For the role of Zn(II), different metal centers other than Zn(II) could be used to study how well they can inhibit the self-cyclization reaction. Again, this would provide testable and measureable insights for testing by experimentalists. The self-cyclization reaction again could be examined in an ancient SerRS lacking the Zn(II) center as well as in CysRS and ThrRS to strengthen its role of inhibiting self cyclization.

6.2 References

- (1) Huang, W. J.; Bushnell, E. A. C.; Francklyn, C. S.; Gault, J. W. *J. Phys. Chem. A* **2011**, *115*, 13050.
- (2) Liu, H. N.; Gault, J. W. *J. Phys. Chem. B* **2008**, *112*, 16874.
- (3) Lv, P. C.; Zhu, H. L. *Curr. Med. Chem.* **2012**, *19*, 3550.
- (4) Zeng, Y.; Kulkarni, A.; Yang, Z.; Patil, P. B.; Zhou, W.; Chi, X.; Van Lanen, S.; Chen, S. *ACS Chemical Biology* **2012**, *7*, 1565.
- (5) Agarwal, V.; Nair, S. K. *Medchemcomm* **2012**, *3*, 887.
- (6) Hurdle, J. G.; O'Neill, A. J.; Chopra, I. *Antimicrob. Agents Chemother.* **2005**, *49*, 4821.

Vita Auctoris

Name: Grant Fortowsky
Place of Birth: Windsor, Ontario, Canada
Year of Birth: 1990

Education:

| | |
|-----------------------|------------------|
| M.Sc Chemsitry | |
| University of Windsor | 2014 |
| | Windsor, Ontario |
| B.Sc Chemistry | |
| University of Windsor | 2012 |
| | Windsor, Ontario |